DUKE MATHEMATICAL JOURNAL

Volume 9, Number 3
SEPTEMBER, 1942

DUKE UNIVERSITY PRESS
DURHAM, N. C.

DUKE MATHEMATICAL JOURNAL

This periodical is published quarterly under the auspices of Duke University by Duke University Press at Durham, North Carolina. It is printed at 8 North Sixth St., Richmond, Virginia by the William Byrd Press.

Entered as second class matter at the Post Office, Durham, North Carolina. Application has been made for additional entry at the Post Office, Richmond, Virginia.

The subscription price for the current year is four dollars, postpaid; back volumes, five dollars each, carriage extra. Subscriptions, orders for back numbers, and notice of change of address should be sent to Duke University Press, Durham, North Carolina.

Individual and institutional members of the Mathematical Association of America may subscribe to the current volume at half price. To get the reduced price, orders for subscriptions must bear the mention "Member MAA". If an order at the reduced price is placed through an agent, the purchaser must pay any commission charge incurred.

Since 1935 the Mathematical Association of America has given the Duke Mathematical Journal an annual subsidy, in return for which the half-rate has been allowed. Having served its purpose of aiding the establishment of the Journal, the subsidy is to be discontinued at the end of 1942. In view of the help already received from the Association, however, Duke University Press will for at least five years continue to allow the half-rate to any one who in 1942 is a subscriber at the reduced rate provided his subscription to the Journal and membership in the Association remain unbroken. This arrangement is expected to be permanent, but the Press reserves the right to modify or withdraw it after the five years and to change the basic rate at the beginning of any calendar year.

Manuscripts and editorial correspondence should be addressed to Duke Mathematical Journal, 4785 Duke Station, Durham, North Carolina. An Author's Manual containing detailed information about the preparation of papers for publication will be sent on request.

Authors are entitled to one hundred free reprints. Additional copies will be supplied at cost. All reprints will be furnished with covers unless the contrary is specifically requested.

The American Mathematical Society is officially represented on the Editorial Board by Professors Murray and Ward.

NON-ANALYTIC CLASS FIELD THEORY AND GRUNWALD'S THEOREM

By George Whaples

Introduction. I present here a new, simple, and non-analytic proof of the theorem of Grunwald [6] on the existence of cyclic fields of minimal degree with given local properties—the theorem which is applied in the proof that every division algebra is cyclic. I also include a systematic exposition of the main part of class field theory, from the index theorems on, in the considerably simplified form made possible by proving the theorems directly in terms of the ideal elements, and the Artin-symbol for ideal elements, introduced by Chevalley [1]. No such account exists in the literature, since the non-analytic proofs recently given by Chevalley [2] are expressed in terms of topology, infinite Abelian extensions, and group characters. This exposition includes non-analytic proofs of the theorem that every class field is Abelian and of the norm index theorem for general fields, a very simple proof of the theorem on possible values of the norm residue symbol, and a proof of the theorem on ramification of class fields (which is omitted in [2]). The proof of this ramification theorem is new, and somewhat shorter than that which would be obtained by the method of Herbrand and Chevalley [3].

1. **Definitions.** Let k be any finite algebraic extension of the rational field. A prime divisor p of k is a symbol associated with a valuation $|\ |_p$ of k. (For the theory of valuations, see [8], [11; X], or [12; III].) The field obtained by adjoining to k all limits of Cauchy sequences (with respect to $|\ |_p$) will be called the p-adic completion of k and will be denoted by the symbol k_p. An ideal element of k (k-idèle) is a vector $\mathfrak{a} = (\alpha_p)$, where p runs through all prime divisors of k, each α_p is a non-zero element of k_p, and α_p is a p-adic unit for all but a finite number of p. Among all prime divisors are, of course, included the infinite prime divisors (those associated with Archimedean valuations). The number α_p is called the p-component of $\mathfrak{a}$. The product of two ideal elements is formed by taking products of their p-components; it is again an ideal element. If K is a finite algebraic extension of k and $\mathfrak{A} = (\mathrm{A}_P)$ is an ideal element of K (P runs through all prime divisors of K), then $N_{K|k}\mathfrak{A}$ is by definition the ideal element of k with p-components

$$\alpha_p = \prod_{P|p} N_{K_P|k_p} \mathrm{A}_P .$$

If σ is an automorphism of K, we define the valuation $|\ |_{P^\sigma}$ of K by the equation $|\ \mathrm{A}^\sigma\ |_{P^\sigma} = |\ \mathrm{A}\ |_P$. σ can be extended to an isomorphism of K_P to K_{P^σ} as follows: if an element A_P of K_P is the limit of a sequence A_ν of elements of K, which con-

Received May 7, 1941.

verges relative to $|\ |_P$, we define A_{P^σ} to be the limit of the sequence A_ν^σ of elements of K^σ relative to $|\ |_{P^\sigma}$.

If $\mathfrak{A} = (A_P)$ is an ideal element of K, $\mathfrak{A}^\sigma$ is defined to be the ideal element of K^σ with components $A_{P^\sigma} = A_P^\sigma$. $N_{K|k}\mathfrak{A}$ can be interpreted as a product of conjugates; if $\mathfrak{a} = (\alpha_p)$ is a k-idèle and $\mathfrak{a}^* = (A_P^*)$ is the K-idèle with $A_P^* = \alpha_p$ for each P which divides p, then under the isomorphism $\mathfrak{a} \leftrightarrow \mathfrak{a}^*$ of k-idèles to a subgroup of the K-idèles, $N_{K|k}\mathfrak{A} \leftrightarrow \prod \mathfrak{A}^\sigma$, where σ runs through all k-isomorphisms of K to subfields of a normal extension of k which contains K.

The following letters are used throughout this paper to stand for general elements of frequently used groups. (The last four are defined later, when first used, and are inserted here merely for convenience.)

$\mathfrak{a}$ = all k-idèles.

α = all non-zero k-elements.

α_p = all non-zero k_p-elements.

ϵ_p = all k_p-units. (At an infinite prime spot, every non-zero element is considered a unit.)

$\mathfrak{a}_S$ = all k-idèles with p-component arbitrary at p in S, p-component any unit for p not in S.

α_S = all k-elements which are p-units at all p not in S, arbitrary at p in S.

$\mathfrak{a}_p$ = all k-idèles with p-component arbitrary, other components 1.

$\mathfrak{e}_p$ = all k-idèles with p-component any unit, other components 1.

$\mathfrak{b}^S$ = all k-idèles with p-component 1 for p in S, p-component any unit for p not in S.

$\mathfrak{a}_{[S]}$ = all k-idèles with p-component arbitrary for p in S, p-component 1 for p not in S.

As usual in class field theory, these letters are sometimes used to stand for the whole group (as in the symbol $(\mathfrak{a} : \mathfrak{b})$, which stands for the index of the group $\mathfrak{b}$ in the group $\mathfrak{a}$) and sometimes to stand for an element of the group. It is always clear from the context which usage is meant. This notation was introduced by Hasse [7; Part Ia, 235]. I will identify a non-zero element α of k with the idèle which has $\alpha_p = \alpha$ for every p. Thus α may be regarded as a subgroup of $\mathfrak{a}$.

Capital letters are used for similar groups in the field K: $\mathfrak{A}$ = all K-idèles, A = all K-elements, E_P = all K_P-units, etc.

To simplify the formulas, "$N_{K|k}$" is sometimes abbreviated to "N", and "$N_{K_P|k_p}$" to "N_P". No other norm symbols are ever abbreviated.

2. **Preliminary computations.** In this section I quote theorems mainly concerned with index computations from [2]. It should be noted that the logical structure of their proof (for which I refer the reader to Chevalley) is much more complicated than that of a single chain of theorems.

LEMMA 1. *There exist sets S, consisting of a finite number of prime divisors of k, such that $\mathfrak{a} = \alpha\mathfrak{a}_S$. If S contains all infinite prime divisors and if s is the number*

of prime divisors in S, then there is a set $\epsilon_1, \epsilon_2, \cdots, \epsilon_s$ of elements of α_S such that ϵ_1 is a root of unity and the others are not, and every α_S is uniquely expressible as a product of rational integral powers of $\epsilon_1, \epsilon_2, \cdots, \epsilon_s$.

LEMMA 2. *If $K_P \mid k_p$ is cyclic, and e_p and f_p are the ramification number and degree of P in $K_P \mid k_p$, then $(\alpha_p : N_P\mathrm{A}_P) = e_p f_p$ and $(\epsilon_p : N_P\mathrm{E}_P) = f_p$.*

LEMMA 3. *If $K \mid k$ is cyclic, if S is a finite set of prime k-divisors which includes all divisors ramified in $K \mid k$ and all infinite divisors, and if $\mathfrak{a} = \alpha\mathfrak{a}_S$, then*

$$(\mathfrak{a} : \alpha N\mathfrak{A}) = (K : k)(\alpha_S \cap N\mathfrak{A}_S : N\mathrm{A}_S).$$

LEMMA 4. *If $K \mid k$ is cyclic of prime power degree, there are an infinite number of p such that $f_p = (K : k)$.*

In the following lemma (as in the entire paper) r stands for the rational field and, if p is a finite k-divisor, $N_{k|r}p$ stands for the number of residue classes modulo p.

LEMMA 5. *If k contains the n-th roots of unity and p^ν is the power of p which divides n exactly, then*

$$(\epsilon_p : \epsilon_p^n) = nN_{k|r}p^\nu \text{ if } p \text{ is finite};$$

$$(\alpha_p : \alpha_p^n) = n^2N_{k|r}p^\nu \text{ if } p \text{ is finite};$$

$$(\alpha_p : \alpha_p^n) = n = 2 \text{ if } p \text{ is infinite and real, and } n = 2;$$

$$(\alpha_p : \alpha_p^n) = 1 \text{ all other cases.}$$

LEMMA 6. *If $K \mid k$ is Abelian, then $(\mathfrak{a} : \alpha N\mathfrak{A}) \leq (K : k)$ and $(\mathfrak{a}_p : N_P\mathfrak{A}_P) \leq (K_P : k_p)$.*

Finally, I quote a lemma not proved in [2]:

LEMMA 7. *If $\mathbf{k}$ is a p-adically closed field which is of finite degree over r_p, and $\mathbf{K} \mid \mathbf{k}$ is of finite degree, then there exist fields K and k, K of finite degree over k, k of finite degree over r, with prime divisors P and p such that $\mathbf{K} \cong K_P$ and $\mathbf{k} \cong k_p$. If $\mathbf{K} \mid \mathbf{k}$ is normal, K and k may be so chosen that $K \mid k$ is normal with Galois group isomorphic to that of $\mathbf{K} \mid \mathbf{k}$.*

Proof. It follows easily from Hensel's irreducibility theorem ([8; 71] or [13; 90-91]) that there exist fields L, l, such that L is of finite degree over l and $L_P = \mathbf{K}$, $l_p = \mathbf{k}$. If $\mathbf{K} \mid \mathbf{k}$ is normal, let K be the smallest extension of L which is normal over l and let k be the decomposition field of p in K; then $K_P = \mathbf{K}$, $k_p = \mathbf{k}$, and the Galois groups are isomorphic.

Lemma 3 contains the second fundamental inequality of class field theory; proofs of it even in the classical theory are non-analytic. Lemma 6 is the first

fundamental inequality; before the work of Chevalley, the proof of this required use of the zeta function. Use of the zeta function gives $(\mathfrak{a} : \alpha N\mathfrak{A}) \leq (K : k)$ for all fields; we will be able to remove our restriction that $K \mid k$ be Abelian very easily after the reciprocity law and the existence theorem have been proved.

An algebraic extension $K \mid k$ will be called a class field (to the group $\alpha N\mathfrak{A}$) when $(\mathfrak{a} : \alpha N\mathfrak{A}) = (K : k)$. From Lemmas 3 and 5 it follows at once that every cyclic field is a class field and that if $K \mid k$ is cyclic, an element of k is a norm (from K to k) if and only if it is everywhere a local norm. To prove that every Abelian field is a class field we must first prove the reciprocity law.

3. **The reciprocity law.** If $K \mid k$ is Abelian and unramified at p, $(K \mid k/p)$ stands for the element τ of the Galois group of $K \mid k$ such that for all integral A in K,

$$\mathrm{A}^{\tau} \equiv \mathrm{A}^{N_{k|rp}} \pmod{P}. \tag{1}$$

It is well known that there is one and only one such τ and that it generates the decomposition group of p and hence is of period f_p. If S is a finite set of prime divisors, which includes all those which are ramified in $K \mid k$, and $\mathfrak{a}$ is a k-idèle whose p-component is a local norm at all p in S, then $[\mathfrak{a}, K \mid k]_S$ is defined by

$$[\mathfrak{a}, K \mid k]_S = \prod (K \mid k/p)^{\nu(p)} \qquad (p \text{ not in } S,\ p \text{ finite}),$$

where $\nu(p)$ is the p-ordinal of the p-component of $\mathfrak{a}$. The product converges because $\nu(p) = 0$ for all but a finite number of p. This definition is a preliminary one; it will later be extended so as to be equivalent to that of Chevalley.

It is easily shown that

$$[\mathfrak{a}, K \mid k]_S \cdot [\mathfrak{b}, K \mid k]_S = [\mathfrak{ab}, K \mid k]_S , \tag{2}$$

provided the symbols are defined. If $\Omega \mid k$ is algebraic and S' is the set of all Ω-divisors $P_{\mathfrak{Q}}$ which divide p in S, then

$$[\mathfrak{A}_{\mathfrak{Q}} , K\Omega \mid \Omega]_{S'} = [N_{\mathfrak{Q}|k}\mathfrak{A}_{\mathfrak{Q}} , K \mid k]_S , \tag{3}$$

since $(K\Omega \mid \Omega/P_{\mathfrak{Q}}) = (K \mid k/N_{\mathfrak{Q}|k}P_{\mathfrak{Q}})$ for every $P_{\mathfrak{Q}}$ not in S'. (This follows easily from (1).) By use of (2), (3), and Lemma 4, it is possible to construct idèles $\mathfrak{a}$ such that $[\mathfrak{a}, K \mid k]_S$ is any given automorphism of $K \mid k$.

For any $\mathfrak{a} = (\alpha_p)$ there is an α such that $[\alpha\mathfrak{a}, K \mid k]_S$ is defined; for we can choose an α such that, for every p in S, $\alpha\alpha_p \equiv 1 \pmod{p^{\nu}}$, with ν so large that $\alpha\alpha_p$ is a local norm. If $\mathfrak{a}$ is a fixed idèle, we define $(\mathfrak{a}, K \mid k)_S$ to be the set of all values taken on by $[\alpha\mathfrak{a}, K \mid k]_S$ as α ranges over all values such that $[\alpha\mathfrak{a}, K \mid k]_S$ is defined. (It will be seen later that $(\mathfrak{a}, K \mid k)_S$ is really single-valued.) If G^* is the subgroup of the Galois group G of $K \mid k$ consisting of all values of $[\alpha, K \mid k]_S$, then if $[\alpha\mathfrak{a}, K \mid k]_S = \tau$ for some α, $(\mathfrak{a}, K \mid k)_S$ is the coset τG^* of G^* in G. So if $\mathfrak{a}^*$ is the set of $\mathfrak{a}$ such that, for some α, $[\alpha\mathfrak{a}, K \mid k]_S = 1$, i.e., the set of $\mathfrak{a}$ such that $(\alpha\mathfrak{a}^*, K \mid k)_S = G^*$, then it follows from (2) that $\mathfrak{a}/\mathfrak{a}^* \cong G/G^*$.

THEOREM 1 (Reciprocity law). *If $K \mid k$ is Abelian and S includes all prime spots at which $K \mid k$ is ramified, then $(\mathfrak{a}, K \mid k)_S$ is a single element of the Galois group of $K \mid k$. $\mathfrak{a} \to (\mathfrak{a}, K \mid k)_S$ thus gives a homorphism of the group of all idèles onto the Galois group. Under this homorphism exactly the elements of the subgroup $\alpha N\mathfrak{A}$ correspond to the identity automorphism. Hence $(\mathfrak{a} : \alpha N\mathfrak{A}) = (K : k)$; every Abelian field is a class field.*

Proof. The above discussion shows that Theorem 1 is equivalent to the conjunction of the statements $\mathfrak{a}^* = \alpha N\mathfrak{A}$, $G^* = (1)$, and $(\mathfrak{a} : \alpha N\mathfrak{A}) = (K : k)$.

Case 1. $K \mid k$ is cyclotomic, generated by a primitive m-th root ζ of 1, and S includes all infinite p and all divisors of m.

It suffices to prove that $G^* = (1)$; for then $(\mathfrak{a} : \mathfrak{a}^*) = (K : k)$, and since $\mathfrak{a}^* \supset \alpha N\mathfrak{A}$ and by Lemma 6 $(\mathfrak{a} : \alpha N\mathfrak{A}) \leq (K : k)$, it follows that $\mathfrak{a}^* = \alpha N\mathfrak{A}$.

If p is finite and prime to m, and ζ is a primitive m-th root of unity, the congruence

$$\zeta^{(K \mid k/p)} \equiv \zeta^{N_{k \mid r} p} \pmod{p}$$

can be sharpened to an equality; for if p is finite and $\zeta^r \equiv \zeta^u \pmod{p}$, where $\zeta^r \neq \zeta^u$, then p divides $1 - \zeta^{r-u}$, so p divides $(1 - \zeta)(1 - \zeta^2) \cdots (1 - \zeta^{m-1}) = f(1) = m$, where $f(x) = (x - \zeta)(x - \zeta^2) \cdots (x - \zeta^{m-1}) = 1 + x + x^2 + \cdots + x^{m-1}$; hence such a p is never prime to m. So whenever $[\alpha, K \mid k]_S$ is defined, we see by (1) that $\zeta^{[\alpha, K \mid k]_S} = \zeta^a$, where a is a rational integer such that $a \equiv \mid N_{k \mid r}\, \alpha \mid \pmod{m}$. ($\mid N_{k \mid r}\alpha \mid$ is the ordinary absolute value of this norm, and is thus equal to the norm of the ideal generated by α, in case α is an integer of k.)

But when $[\alpha, K \mid k]_S$ is defined, α is a local norm at each p in S, so there is an A in K such that $\alpha N\mathrm{A} \equiv 1 \pmod{m}$ and is positive at each infinite prime divisor of k. (The existence of A depends on a generalization of the Chinese remainder theorem which is best stated as an approximation theorem: *If a finite set S of prime k-divisors is given (infinite p not excluded), an element α_p of k_p is chosen for each p in S, and ϵ is any positive number, then there is an α in k such that $\mid \alpha - \alpha_p \mid_p < \epsilon$ for each p in S.* This theorem will be used repeatedly.) Then $N_{k \mid r}\,(\alpha N_{K \mid k}\mathrm{A}) \equiv 1 \pmod{m}$ and is positive, so $[\alpha N\mathrm{A}, K \mid k]_S = 1$. But $[N\mathrm{A}, K \mid k]_S = 1$ always, because $(K \mid k/NP) = (K \mid k/p)^f$, where P divides p and f is the degree of P, so $(K \mid k/NP) = 1$ for all P. Hence $[\alpha, K \mid k]_S = 1$, and Theorem 1 is proved for Case 1.

Case 2. $K \mid k$ is cyclic and S contains all primes ramified in $K \mid k$.

In this case it suffices to prove that $\mathfrak{a}^* = \alpha N\mathfrak{A}$; for since $K \mid k$ is cyclic, we know by Lemmas 3 and 6 that $(\mathfrak{a} : \alpha N\mathfrak{A}) = (K : k)$, and since $(G : G^*) = (\mathfrak{a} : \mathfrak{a}^*)$, $\mathfrak{a}^* = \alpha N\mathfrak{A}$ implies that $G^* = (1)$. Also, $\mathfrak{a}^* \supset \alpha N\mathfrak{A}$ is already known. Assume, therefore, that a fixed idèle $\mathfrak{a}^*$ is given such that $[\mathfrak{a}^*, K \mid k]_S = 1$; we shall prove that $\mathfrak{a}^*$ is in the group $\alpha N\mathfrak{A}$.

Since $[\mathfrak{a}^*, K \mid k]_S$ is defined, $\mathfrak{a}^*$ is a local norm at all p in S. Furthermore, at all infinite p outside S and all finite p outside S at which $\mathfrak{a}^*$ is a local unit, $\mathfrak{a}^*$ is a local norm by Lemma 2. So for suitably chosen $\mathfrak{A}$, $\mathfrak{a}^* N\mathfrak{A}$ has p-component

1 at all p in S, all infinite p, and all finite p outside S at which $\mathfrak{a}^*$ is a local unit. Thus we will assume that $\mathfrak{a}^*$ has p-component 1 at all p outside T, where T is a finite set of finite prime divisors not in S. (This element $\mathfrak{a}^*$ remains fixed throughout the rest of this proof.)

Choose a field K' such that

$K' \cap K = k$;

$K' \mid k$ is generated by a primitive (m')-th root of unity, where m' is prime to all p whose a^*-ordinal is not 0, i.e., to all p in T;

the Galois group of $K' \mid k$ contains an element σ' of period divisible by $n = (K : k)$;

and for any given p in T, choose a field K'' such that

$K'' \cap K'K = k$;

K'' is generated by a primitive (m'')-th root of unity, where m'' is prime to all p in T;

$(K'' \mid k/p)$ is of period divisible by n.

The existence of fields with these properties follows from a theorem proved non-analytically in [10]. In fact, only the first and simplest of the theorems proved in that paper is needed here; this is an advantage of this proof of the reciprocity law over the proof which uses the theory of algebras.

Since the intersections by pairs of K, K', and K'' are all equal to k, the Galois group of $KK'K'' \mid k$ is the direct product of the groups of the individual fields over k; we will interpret an element of the Galois group of $K \mid k$ as the automorphism of $KK'K'' \mid k$ which produces the given automorphism of $K \mid k$ and leaves the other two fields invariant, and will interpret elements of the Galois groups of $K' \mid k$ and $K'' \mid k$ similarly.

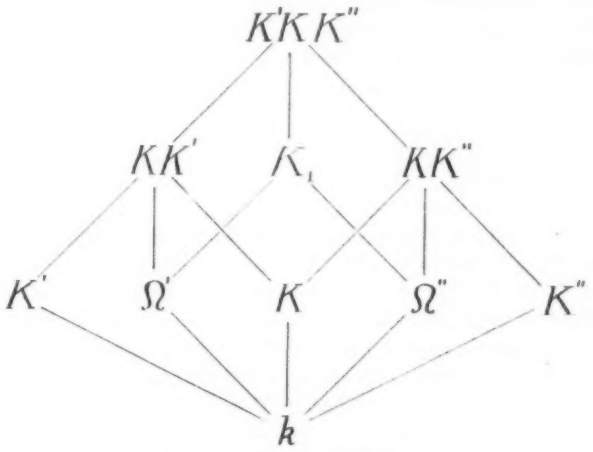
Let Ω' be the subfield of KK' left invariant by $\langle \sigma\sigma' \rangle$, where σ is the generator of the Galois group of $K \mid k$. ($\langle \sigma\sigma' \rangle$ stands for the cyclic group generated by the element $\sigma\sigma'$.) Let Ω'' be the subfield of KK'' left invariant by $\langle \sigma^r\sigma'' \rangle$, where $(K \mid k/p) = \sigma^r$. In $K'K$, $K'\Omega'$ is left invariant by $\langle \sigma \rangle \cap \langle \sigma\sigma' \rangle = (1)$; hence $K'\Omega' = K'K$. Similarly, $K''\Omega'' = K''K$. $KK'' \mid \Omega'$ and $KK'' \mid \Omega''$ are thus cyclotomic, and the results of Case 1 apply to them. Let K_1 be the subfield of $KK'K''$ left invariant by $\langle \sigma^r\sigma''\sigma'' \rangle$. K_1 includes Ω' and Ω''.

The structure of the set of fields which we have constructed is given by Fig. 1. We now want to compute with bracket symbols, defined in these fields, using the rules (2) and (3). All the bracket symbols used from now on are to be understood as defined relative to the set S'' of all prime divisors of the field concerned which are infinite, or divide primes in S, or divide $m'm''$. Then S'' contains all possible divisors of primes in S and also contains all primes ramified in any of the fields in our diagram. Thus (3) is valid for any three fields of the diagram to which it could be applied.

Let $\mathfrak{T}'$ be an Ω'-idèle such that

$$[\mathfrak{T}', KK' \mid \Omega'] = \sigma\sigma', \tag{4}$$

and let $\mathfrak{d} = N_{\Omega'|k}\mathfrak{D}'$. The method of proof is to show, by means of the norm relation (3), that every k-idèle for which the bracket symbol is defined is congruent to a power of $\mathfrak{d}$ (mod $\alpha N_{K|k}\mathfrak{A}$). Though this fact is not used in the proof, it follows from (3) that $[\mathfrak{d}, KK' \mid k] = \sigma\sigma'$, so that $[\mathfrak{d}, K \mid k] = \sigma$.



FIELDS USED IN PROOF OF THEOREM 1

Let $\mathfrak{B}_1$ be a K_1-idèle such that

$$[\mathfrak{B}_1 , KK'K'' \mid K_1] = \sigma^r\sigma'^r\sigma''.$$

Then if $\mathfrak{B}' = N_{K_1|\Omega'}\mathfrak{B}_1$, $[\mathfrak{B}', KK'K'' \mid \Omega'] = \sigma^r\sigma'^r\sigma''$; from our definitions of σ, σ', and σ'', it follows that

$$[\mathfrak{B}' , KK' \mid \Omega'] = (\sigma\sigma')^r.$$

Similarly, if $\mathfrak{B}'' = N_{K_1|\Omega''}\mathfrak{B}_1$,

$$[\mathfrak{B}'', KK'' \mid \Omega''] = \sigma^r\sigma''. \tag{5}$$

Since $KK' \mid \Omega'$ is cyclotomic, $\mathfrak{D}'' \equiv \mathfrak{B}'$ (mod $A_{\Omega'} N_{KK'|\Omega'}\mathfrak{A}_{KK'}$); so if $\mathfrak{b} = N_{K_1|k}\mathfrak{B}_1$,

$$\mathfrak{d}^r \equiv \mathfrak{b} \pmod{\alpha N_{K|k}\mathfrak{A}}; \tag{6}$$

for $\mathfrak{d}^r\mathfrak{b}^{-1}$ is in $N_{\Omega'|k}(A_{\Omega'}N_{KK'|\Omega'}\mathfrak{A}_{KK'}) \subset \alpha N_{KK'|k}\mathfrak{A}_{KK'} \subset \alpha N_{K|k}\mathfrak{A}$.

Now $(K'' \mid k/p) = \sigma''$ and $(K \mid k/p) = \sigma^r$; so $(KK'' \mid k/p) = \sigma^r\sigma''$ and Ω'' is exactly the decomposition field of p in KK'', i.e., the largest subfield in which p splits into prime factors of degree 1. Hence there is a prime divisor P'' in Ω'' such that $N_{\Omega''|k}P'' = p$. If $\mathfrak{R}''$ is any Ω''-idèle which is of ordinal 1 at P'' and is equal to 1 everywhere else, and if $\mathfrak{k} = N_{\Omega''|k}\mathfrak{R}''$, then $[\mathfrak{R}'', KK'' \mid \Omega''] = (KK'' \mid k/p) = \sigma^r\sigma''$ and

$$[\mathfrak{k}, K \mid k] = \sigma^r. \tag{7}$$

Since $KK'' \mid \Omega''$ is cyclotomic, (5) shows that $\mathfrak{R}'' \equiv \mathfrak{B}''$ (mod $A_{\Omega''} N_{KK'' \mid \Omega''} \mathfrak{A}_{KK''}$); just as above, then, we get

$$\mathfrak{f} \equiv \mathfrak{b}^r \pmod{\alpha N_{K \mid k} \mathfrak{A}}. \tag{8}$$

Let $\mathfrak{a}^*$ be of p-ordinal t; if $\mathfrak{a}_p^*$ is the idèle whose p-component is equal to the p-component of $\mathfrak{a}^*$ and whose other components are 1, then $\mathfrak{a}_p^*$ differs from $\mathfrak{f}^t$ by at most a p-unit, and since p is unramified in $K \mid k$, every p-unit is a norm in $K_P \mid k_p$. So by use of (8) and (7), we see that

$$\mathfrak{a}_p^* \equiv \mathfrak{b}^{rt} \pmod{\alpha N_{K \mid k} \mathfrak{A}}, \tag{9}$$

where $[\mathfrak{a}_p^*, K \mid k] = \sigma^{rt}$.

The same argument can be repeated for each p in T, K', $\mathfrak{D}'$, and $\mathfrak{b}$ being kept constant, but K'' changed for each p. By multiplying the equations of type (9), we find that $\mathfrak{a}^* \equiv \mathfrak{b}^m \pmod{\alpha N_{K \mid k} \mathfrak{A}}$, where m is an integer such that $[\mathfrak{a}^*, K \mid k] = \sigma^m$. But since $[\mathfrak{a}^*, K \mid k] = 1$, n divides m; so $\mathfrak{b}^m$ is an n-th power and is thus in $N_{K \mid k} \mathfrak{A}$. Hence $\mathfrak{a}^*$ is in $\alpha N_{K \mid k} \mathfrak{A}$, and the theorem is proved for Case 2.

To prove the theorem in general, we need only note that any Abelian field $K \mid k$ is a product of cyclic fields. If α is any element of k for which the symbol is defined, $[\alpha, K \mid k]$ leaves every cyclic subfield of $K \mid k$ invariant; hence $[\alpha, K \mid k] = 1$, $G^* = (1)$, and the theorem follows by the argument used at the beginning of Case 1.

If $S_1 \supset S$, then $(\mathfrak{a}, K \mid k)_{S_1} = (\mathfrak{a}, K \mid k)_S$ since, for some suitable α, $\alpha\mathfrak{a}$ is a local norm at all p outside S_1 ; hence the primes in $S_1 - S$ contribute nothing to $[\alpha\mathfrak{a}, K \mid k]_S$; hence $[\alpha\mathfrak{a}, K \mid k]_S = [\alpha\mathfrak{a}, K \mid k]_{S_1}$. The subscript S will be dropped from now on.

4. **The existence theorem.** If S is a finite set of k-divisors, $\mathfrak{b}^S$ will stand for the set of all idèles which are 1 at all p in S and are p-units at all other p. An idèle group $\mathfrak{h}$ is called *admissible* if it contains $\mathfrak{b}^S$ for some finite S, and is of finite index in $\mathfrak{a}$. If h contains the group $\mathfrak{e}_p$ of all idèles which are units at p, and are 1 elsewhere, we say that $\mathfrak{h}$ is *unramified* at p. If h contains the group $\mathfrak{a}_p$ of all idèles which are arbitrary at p, and are 1 elsewhere, we say that $\mathfrak{h}$ *splits completely* at p. We make these definitions in the hope of proving that every admissible $\mathfrak{h}$ has a class field, and that this class field is unramified (splits completely) at exactly the p at which $\mathfrak{h}$ is unramified (splits completely).

The following lemma shows that we can hope to prove the existence of a class field to $\mathfrak{h}$ only when $\mathfrak{h}$ is admissible.

LEMMA 8. *If $K \mid k$ is of finite degree, $\alpha N\mathfrak{A}$ is admissible and is unramified at every p which is unramified in $K \mid k$.*

Proof. Let $\Omega \mid k$ be the smallest normal extension containing K. It follows from the Hilbert theory of the structure of the decomposition group that the Galois group of $\Omega_P \mid k_p$ is solvable for each P [12; Theorems III 10, A, B, and C].

Hence repeated application of Lemma 2 shows that $(\alpha_p : N_P A_P)$ is finite and that if p is unramified in $K \mid k$, $(\epsilon_p : N_P E_P) = 1$. For these statements are true if we substitute Ω for K; but the groups of norms from K to k contain the corresponding groups of norms from Ω to k, and if p is unramified in $K \mid k$, p is also unramified in $\Omega \mid k$. So if S contains all p ramified in $K \mid k$, $\mathfrak{b}^S$ is contained in $\alpha N\mathfrak{A}$. If S is chosen so large that $\mathfrak{a} = \alpha\mathfrak{a}_S$, then $(\mathfrak{a} : \alpha N\mathfrak{A}) = (\alpha\mathfrak{a}_S : \alpha N\mathfrak{A}_{S'}) \leq (\mathfrak{a}_S : N\mathfrak{A}_{S''})$; this last index is the product over all p in S of the indices $(\alpha_p : N_P A_P)$ and is therefore finite. Hence $(\mathfrak{a} : \alpha N\mathfrak{A})$ is finite.

LEMMA 9. *If the idèle group $\mathfrak{b}$ possesses an Abelian class field $K \mid k$, and if Ω is any field such that $\alpha N_{\Omega|k}\mathfrak{A}_\Omega \subset \mathfrak{b}$ (where Ω and K are both subfields of some larger field), then $\Omega \supset K$.*

For, $K\Omega \mid \Omega$ is Abelian, so the reciprocity law holds and

$$(\mathfrak{A}_\Omega , K\Omega \mid \Omega) = (N_{\Omega|k}\mathfrak{A}_\Omega , K \mid k).$$

But by assumption, $K \mid k$ is a class field to $\mathfrak{b}$ and $N_{\Omega|k}\mathfrak{A}_\Omega \subset \mathfrak{b} = \alpha N_{K|k}\mathfrak{A}$; hence $(\mathfrak{A}_\Omega , K\Omega \mid \Omega) = 1$ for any $\mathfrak{A}_\Omega$. But this symbol takes on as values all elements of the Galois group of $K\Omega \mid \Omega$; hence $K\Omega = \Omega$, $K \subset \Omega$.

LEMMA 10. *Let $\mathfrak{b}$ be an admissible group of k-idèles and let Ω be a cyclic extension of k such that the group $\mathfrak{H}$ of all Ω-idèles whose $N_{\Omega|k}$ are in $\mathfrak{b}$ has an Abelian class field C. Then $\mathfrak{b}$ has an Abelian class field over k, and this class field is a subfield of C.*

Proof. Let the Abelian class field C to $\mathfrak{H}$ be contained in the field C', which is normal over k. If θ is any isomorphism of C to another subfield of C', θ takes Ω into itself since Ω is normal over k; hence C^θ is class field to $\mathfrak{H}^\theta = \mathfrak{H}$, so by Lemma 9, $C^\theta = C$. So C is normal over k. Let τ be an automorphism of C which leaves every element of Ω invariant; let $\tau = (\mathfrak{A}, C \mid \Omega)$. Then if σ is an automorphism of $C \mid k$ which generates the Galois group of $\Omega \mid k$, $(\mathfrak{A}^\sigma, C \mid \Omega) = \sigma\,(\mathfrak{A}, C \mid \Omega)\,\sigma^{-1} = \sigma\tau\sigma^{-1}$. Since $N_{\Omega|k}\,\mathfrak{A}^{1-\sigma} = 1$ and is therefore in $\mathfrak{b}$, $\mathfrak{A}^{1-\sigma}$ is by construction in $\mathfrak{H}$; so $(\mathfrak{A}^\sigma, C \mid \Omega) = (\mathfrak{A}, C \mid \Omega) : \sigma\tau = \tau\sigma$. But any automorphism of $C \mid k$ is of form $\sigma^r\tau$, where τ is an automorphism leaving Ω invariant. Since σ is commutative with any τ and since, by our assumption that $C \mid \Omega$ is Abelian, any two τ's are commutative with each other, $C \mid k$ is Abelian.

So $C \mid k$ is class field to $\alpha N_{C|k}\,\mathfrak{A}$, which is a subgroup of $\mathfrak{b}$. If K is the subfield of C left invariant by the group of all automorphisms $(\mathfrak{b}, C \mid k)$, then K is a class field to $\mathfrak{b}$ itself. For since $(\mathfrak{a}, C \mid k)$ has the same effect on K as has $(\mathfrak{a}, K \mid k)$, it follows that $(\mathfrak{a}, K \mid k) = 1$ if and only if $\mathfrak{a}$ is in $\mathfrak{b} \cdot \alpha N_{C|k}\,\mathfrak{A} = \mathfrak{b}$. Hence, by the reciprocity law, $\mathfrak{b} = \alpha N_{K|k}\mathfrak{A}$. The argument used in constructing K also proves

LEMMA 11. *If any idèle group $\mathfrak{b}$ has an Abelian class field K and $\mathfrak{a} \supset \mathfrak{b}_1 \supset \mathfrak{b}$ and if K_1 is the field of all elements of K left invariant by the group $(\mathfrak{b}_1 , K \mid k)$, then K_1 is class field to $\mathfrak{b}_1$.*

THEOREM 2. *If $\mathfrak{b}$ is any admissible group of k-idèles, then $\mathfrak{b}$ has an Abelian class field.*

Proof. *Case* 1. $\mathfrak{a}/\mathfrak{h}$ is of type $l, l, \cdots, l$, where l is prime, and k contains the l-th roots of unity.

$\mathfrak{h} \supset \mathfrak{a}^l$, and since $\mathfrak{h}$ is admissible, we can choose a finite set S such that $\mathfrak{h} \supset \alpha\mathfrak{a}^l\mathfrak{b}^S$. Choose S so that S also contains all divisors of l and all infinite p, and $\mathfrak{a} = \alpha\mathfrak{a}_S$. Then it follows that

$$(\mathfrak{a} : \alpha\mathfrak{a}^l\mathfrak{b}^S) = (\alpha\mathfrak{a}_S : \alpha\mathfrak{a}_S^l\mathfrak{b}^S) = \frac{(\mathfrak{a}_S : \mathfrak{a}_S^l\mathfrak{b}^S)}{(\alpha \cap \mathfrak{a}_S : \alpha \cap \mathfrak{a}_S^l\mathfrak{b}^S)}$$

(10)

$$= \frac{\prod_{p \text{ in } S} (\alpha_p : \alpha_p^l)}{(\alpha_S : \alpha_S^l)} = \frac{l^{2r_1+r_1+r_1+2r_2}}{l^r} = l^r,$$

where s_1 is the number of finite prime spots in S, r_1 the number of real prime spots of k, and r_2 the number of complex prime spots of K, so that $s = s_1 + r_1 + r_2$ and $(k : r) = r_1 + 2r_2$. The first of the equations (10) is obvious. We get the second by applying the principle that if σ is a homomorphism of a group g and $(g : h)$ is finite, then if g_1 and h_1 are the subgroups which are mapped onto (1) by σ, $(g^\sigma : h^\sigma) = (g : h)/(g_1 : h_1)$. (Here is the proof of the principle: g/h is homomorphic to g^σ/h^σ, and the subgroup of g/h which goes into 1 under this homomorphism is $g_1h^\sigma/h^\sigma \cong g_1/h_1$, so that $(g : h) = (g^\sigma : h^\sigma)(g_1 : h_1)$.) Take g and h to be $\mathfrak{a}_S$ and $\mathfrak{a}_S^l\mathfrak{b}^S$, and σ to be the homomorphism mapping each element of $\mathfrak{a}_S$ onto its residue class in $\alpha\mathfrak{a}_S/\alpha$. Clearly $(\alpha\mathfrak{a}_S : \alpha\mathfrak{a}_S^l\mathfrak{b}^S)$ is equal to the index $((\alpha\mathfrak{a}_S/\alpha) : (\alpha\mathfrak{a}_S^l\mathfrak{b}^S/\alpha))$ of the corresponding groups of residue classes; the latter index is, by our principle, equal to the third expression in (10). To get the third equation we use the fact that $\alpha \cap \mathfrak{a}_S^l\mathfrak{b}^S = \alpha_S^l$. This is a weakened form of a lemma proved in [2; 411]. The rest follows easily by use of Lemmas 5 and 1.

Now if $K \mid k$ is the largest extension which is unramified except at primes in S and is Abelian of type $l, l, \cdots, l$ over k, then, by the theory of Kummer fields (for a brief exposition of this theory, see [13]), $(K : k) = (\alpha_S\alpha^l : \alpha^l) = l^r$. By Theorem 1 and Lemma 8, $K \mid k$ is a class field to a group including $\alpha\mathfrak{a}^l\mathfrak{b}^S$; since, by (10), $(\mathfrak{a} : \alpha\mathfrak{a}^l\mathfrak{b}^S) = (K : k)$, $K \mid k$ is a class field for exactly $\alpha\mathfrak{a}^l\mathfrak{b}^S$. Since $\mathfrak{h} \supset \alpha\mathfrak{a}^l\mathfrak{b}^S$, K contains a subfield which is a class field to $\mathfrak{h}$.

Case 2. $\mathfrak{a}/\mathfrak{h}$ is cyclic of prime order l, k arbitrary.

Let k' be the field obtained by adjoining to k a primitive l-th root of unity; then $k' \mid k$ is cyclic. If $\mathfrak{h}'$ is the set of all k'-idèles such that $N_{k'|k}\mathfrak{h}' \supset \mathfrak{h}$, then $\mathfrak{a}'/\mathfrak{h}'$ is cyclic; for since $N_{k'|k}\mathfrak{a}' = 1$ implies $\mathfrak{a}' \subset \mathfrak{h}'$,

$$\mathfrak{a}'/\mathfrak{h}' \cong N_{k'|k}\mathfrak{a}'/N_{k'|k}\mathfrak{h}' = N_{k'|k}\mathfrak{a}'/\mathfrak{h} \cap N_{k'|k}\mathfrak{a}' \cong \mathfrak{h}N_{k'|k}\mathfrak{a}'/\mathfrak{h} \subset \mathfrak{a}/\mathfrak{h}.$$

So $\mathfrak{h}'$ has a class field by Case 1; by Lemma 10, $\mathfrak{h}$ has a class field.

Case 3. $\mathfrak{a}/\mathfrak{h}$ is cyclic of order l^n, k arbitrary.

By Case 2, any $\mathfrak{h}$ has a class field when $\mathfrak{a}/\mathfrak{h}$ is cyclic of order l; suppose it has been proved that $\mathfrak{h}$ has a class field whenever $\mathfrak{a}/\mathfrak{h}$ is cyclic of order l^{n-1}. Then if $\mathfrak{a}$ is generated by $\mathfrak{a}_0 \pmod{\mathfrak{h}}$, the group $\mathfrak{h}''$ generated by $\mathfrak{a}_0^{l^{n-1}}$ and $\mathfrak{h}$ has a class

field K' over k, since $(\mathfrak{a} : \mathfrak{h}') = l^{n-1}$. The group $\mathfrak{H}'$ of all K'-idèles such that $N_{K'|k}\mathfrak{H}' \subset \mathfrak{h}$ has a class field K over k; for $(\mathfrak{A}' : \mathfrak{H}') = (N_{K'|k}\mathfrak{A}' : N_{K'|k}\mathfrak{H}')$ by the mapping principle used in Case 1; and since $N_{K'|k}\mathfrak{A}' = \mathfrak{h}'$ and $N_{K'|k}\mathfrak{H}' = N_{K'|k}\mathfrak{A}' \cap \mathfrak{h} = \mathfrak{h}$, the latter index equals l. By Lemma 10, $\mathfrak{h}$ has a class field over k.

Case 4. $\mathfrak{a}/\mathfrak{h}$ is general Abelian. Then there exist groups $\mathfrak{h}_i$ such that $\mathfrak{a}/\mathfrak{h}_i$ are cyclic of prime power order and $\mathfrak{h} = \bigcap \mathfrak{h}_i$. It follows from the reciprocity law that the product of the class fields for the $\mathfrak{h}_i$ is the class field for their intersection.

COROLLARY 2.1. *A given admissible idèle group has one and only one class field (if we restrict ourselves to subfields of a fixed algebraically closed field) and that class field is Abelian.*

This follows at once from Theorem 2 and Lemma 9.

COROLLARY 2.2. *If $K \mid k$ is any algebraic extension, then $(\mathfrak{a} : \alpha N\mathfrak{A})$ is equal to the degree over k of the largest subfield of K which is Abelian over k.*

For the class field to $\alpha N\mathfrak{A}$ is by Lemma 9 a subfield of K.

The proof of Theorem 2 is only slightly rearranged from that in [2]. It must be supplemented by a proof that if $\mathfrak{h}$ is unramified at p, then its class field is unramified at p; the proof of Theorem 2 gives this when p does not divide the order of $\mathfrak{a}/\mathfrak{h}$, but the remaining case is left in doubt, and its difficulties are not trivial. It is easy to take care of the ramification by means of the method of Herbrand and Chevalley, and in fact the index computations of that method can be much simplified by use of Chevalley's theorems that $(\mathfrak{a} : \alpha\mathfrak{a}^l\mathfrak{b}^S) = l^s$ and $\alpha \cap \mathfrak{a}_S^l\mathfrak{b}^S = \alpha_S^l$ for suitable S, but I prefer to give here a new proof.

Assume, then, that there does exist an Abelian field $K \mid k$, ramified at p, for which $\mathfrak{h} = \alpha N\mathfrak{A}$ is unramified at p; we will derive a contradiction from this. It is convenient to begin by reducing $K \mid k$ to a simple form. First, if $K \mid k$ is not of prime degree, K contains a subfield k' such that $K \mid k'$ is of prime degree l and is ramified at a divisor p' of p. By relation (3), $Kk' \mid k'$ is class field to the group $\mathfrak{h}'$ of all k'-idèles whose norms are in $\mathfrak{h}$; $\mathfrak{h}'$ is also unramified at p'. Since l is prime, the ramification number of p' in $K \mid k'$ is also l. Next, if k' does not contain the l-th roots of unity, let k'' be formed by adjoining them to k', and let p'' be a prime k''-divisor dividing p'. Since the ramification number of p' in $Kk' \mid k'$ is l and the degree of $k'' \mid k'$ is prime to l, the ramification number of p'' in $Kk'' \mid k''$ is l. $Kk'' \mid k''$ is of degree l, and, as before, $\mathfrak{h}''$ is unramified at p''.

So we need only prove that it is impossible to have a field $K \mid k$ which is cyclic of prime degree l, where k contains the l-th roots of unity and a prime p is ramified in $K \mid k$ but unramified in $\alpha N\mathfrak{A}$. If $K \mid k$ is such a field, p is the l-th power of a K-divisor P; if $\mathfrak{B}$ has ordinal 1 at P, and equals 1 everywhere else, then $N\mathfrak{B}$ has ordinal 1 at p and is 1 everywhere else. $N\mathfrak{B}$, together with the group $\mathfrak{e}_p$, thus generates $\mathfrak{a}_p$; since $\alpha N\mathfrak{A} \supset \mathfrak{e}_p$ by assumption, $\alpha N\mathfrak{A} \supset \mathfrak{a}_p$. If S contains all prime k-divisors ramified in $K \mid k$ (including p itself), then $\alpha N\mathfrak{A} \supset \alpha\mathfrak{a}^l\mathfrak{b}^S\mathfrak{a}_p$; hence $K \mid k$

is subfield to the class field to $\alpha\mathfrak{a}^l\mathfrak{b}^S\mathfrak{a}_p$. So if we can show that, for some sufficiently large S, the class field to $\alpha\mathfrak{a}^l\mathfrak{b}^S\mathfrak{a}_p$ is unramified at p, we have the desired contradiction.

Let C be the class field to $\alpha\mathfrak{a}^l\mathfrak{b}^S$, Z the decomposition field of p in C, and C' the class field to $\alpha\mathfrak{a}^l\mathfrak{b}^S\mathfrak{a}_p$. Then since $\alpha N_{Z|k}\mathfrak{A}_Z \supset \alpha\mathfrak{a}^l\mathfrak{b}^S\mathfrak{a}_p \supset \alpha\mathfrak{a}^l\mathfrak{b}^S$, $Z \subset C' \subset C$. Now

$$(C : C') = \frac{(C : k)}{(C' : k)} = \frac{(\mathfrak{a} : \alpha\mathfrak{a}^l\mathfrak{b}^S)}{(\mathfrak{a} : \alpha\mathfrak{a}^l\mathfrak{b}^S\mathfrak{a}_p)} = (\alpha\mathfrak{a}^l\mathfrak{b}^S\mathfrak{a}_p : \alpha\mathfrak{a}^l\mathfrak{b}^S)$$

(11)

$$= \frac{(\mathfrak{a}_p : \mathfrak{a}_p^l)}{(\mathfrak{a}_p \cap \alpha\mathfrak{a}^l\mathfrak{b}^S : \mathfrak{a}_p^l)} = \frac{(\alpha_p : \alpha_p^l)}{(\mathfrak{a}_p \cap \alpha\mathfrak{a}^l\mathfrak{b}^S : \mathfrak{a}_p^l)}.$$

(The fourth equation follows by mapping the groups $\mathfrak{a}_p$ and $\mathfrak{a}_p^l$ into the corresponding groups of cosets in $\mathfrak{a}_p\alpha\mathfrak{a}^l\mathfrak{b}^S/\alpha\mathfrak{a}^l\mathfrak{b}^S$.) $(C : Z) = (C_P : k_p) \leq (\alpha_p : \alpha_p)$, since by the theory of Kummer fields the last index is the degree of the greatest possible Abelian extension of k_p in which every automorphism has period n. Suppose we can find S such that $\mathfrak{a}_p \cap \alpha\mathfrak{a}^l\mathfrak{b}^S = \mathfrak{a}_p^l$. Then by (11), $(C : C') = (\alpha_p : \alpha_p)$, so $(C : Z) \leq (C : C')$. Since $Z \subset C'$, this implies $Z = C'$, so $C' \mid k$ is unramified at p and we have our contradiction.

The following lemma, which is not more difficult to prove than the more restricted one which would suffice, proves the existence of such sets S.

LEMMA 12. *If k is any number field, n any rational integer, S a finite set of prime divisors, and $\mathfrak{a}_{[S]}$ the set of all k-idèles which are arbitrary at p in S and are 1 elsewhere, then there exists a finite set T of prime divisors, such that the elements of T are outside any given finite set U, and*

$$\mathfrak{a}_{[S]} \cap \alpha\mathfrak{a}^n\mathfrak{b}^{S+T} = \mathfrak{a}_{[S]}^n. \tag{12}$$

(To apply this lemma to the ramification theory, let S contain only the prime p; then $\mathfrak{a}_{[S]} = \mathfrak{a}_p$. We must also demand that T include all primes ramified in $K \mid k$; but obviously we do not lose the property (12) by adding extra primes to T.)

Proof. It suffices to show that for properly chosen T the equation $\alpha = \mathfrak{a}_{[S]}\mathfrak{b}^{S+T}\mathfrak{a}^n$ is impossible whenever $\mathfrak{a}_{[S]}$ is not an n-th power. Suppose there exists a particular number α_0 such that $\alpha_0 = \mathfrak{a}_{[S]}\mathfrak{a}^n\mathfrak{b}^{S+T}$, where the idèle $\mathfrak{a}_{[S]}$ of this equation is not an n-th power. Then, since α_0 fails to be an n-th power at some p in S, the field $K_0 = k(\alpha_0^{1/n}) \neq k$, so by an easy consequence of Lemma 4 there is a prime p_0, outside U, which does not split completely in $K_0 \mid k$. Since for all p outside S the p-ordinal of α_0 is divisible by n, $K_0 \mid k$ is unramified except possibly at divisors which are in S, are infinite, or divide n. We take p_0 as the first prime of our set T.

If we can find α_1 such that $\alpha_1 = \mathfrak{a}_{[S]}\mathfrak{a}^n\mathfrak{b}^{S+p_1}$, where $\mathfrak{a}_{[S]}$ is not an n-th power, and let $K_1 = k(\alpha_1^{1/n})$, then as before there is a p_1 outside U which does not split

completely in $K_1 \mid k$. But, since α_1 is an n-th power at p_0, p_0 splits completely in $K_1 \mid k$. If $\alpha_2 = \mathfrak{a}_{[S]}\mathfrak{b}^{S+p_0+p_1}$, where $\mathfrak{a}_{[S]}$ is not an n-th power, the corresponding K_2, and some p_2, have properties similar to K_0, K_1, p_0, p_1, except that both p_0 and p_1 split completely in K_2. We can continue in a similar manner, getting a set of fields $K_0, K_1, \cdots, K_\nu, \cdots$ and a set of divisors $p_0, p_1, \cdots, p_\nu, \cdots$ unless at some stage the equation $\alpha_{\nu+1} = \mathfrak{a}_{[S]}\mathfrak{a}^n\mathfrak{b}^{S+p_0+\cdots+p_\nu}$ has no solutions in which $\mathfrak{a}_{[S]}$ is not an n-th power. This must happen for some ν; for since each p_i splits completely in $K_{i+1}, K_{i+2}, \cdots$, but not in K_i, $k \neq K_\nu \neq K_\nu K_{\nu-1} \neq K_\nu K_{\nu-1}K_{\nu-2} \neq \cdots \neq K_\nu K_{\nu-1} \cdots K_1K_0$. Hence if $K_\nu = K_\nu K_{\nu-1} \cdots K_1K_0$, $(K_\nu : k) \geq l^\nu$, where l is the smallest prime factor of n; but $(K_\nu : k)$ is bounded since if k' is the field obtained by adjoining all n-th roots of unity to k, $K_\nu k' \mid k'$ is Abelian, with each automorphism of period which divides n, and is unramified except at primes which are in S, are infinite, or divide n.

So we have proved Lemma 12 and

THEOREM 3. *If $K \mid k$ is the class field to $\mathfrak{h}$, then $K \mid k$ is ramified at p if and only if $\mathfrak{h}$ is ramified at p, and splits completely at p if and only if $\mathfrak{h}$ splits completely at p.*

The part of this theorem which concerns complete splitting follows at once from the reciprocity law, once the ramification is settled.

This method is in a way rougher than that of Chevalley and Herbrand [3]; the latter method begins with any S which contains all divisors of n and all infinite divisors, and, considering all divisors in S at once, proves that a certain subfield of C is unramified at the proper places, without determining the degree of $C_P \mid k_p$. Here, on the other hand, the method is to produce a C such that the degree of $C_P \mid k_p$ is the largest conceivable.

5. **Local class field theory and the norm residue symbol.** If $K \mid k$ is Abelian, p is a prime divisor in k, and α_p is any non-zero element of k_p, then the norm residue symbol $(\alpha_p, K \mid k/p)$ is defined to be $(\mathfrak{a}_0, K \mid k)^{-1}$, where $\mathfrak{a}_0$ is the idèle whose p-component is α_p and whose other components are 1. I take the inverse to make this definition coincide with that of Hasse [7; II, 25].

It follows at once that $(\mathfrak{a}, K \mid k) = \prod_p (\alpha_p, K \mid k/p)^{-1}$, where α_p are the p-components of $\mathfrak{a}$ (Chevalley's definition [1] of $(\mathfrak{a}, K \mid k)$). This product is meaningful because its factors are 1 for all but a finite number of prime spots. Also, it is evident that, if α is in k, $\prod_p (\alpha, K \mid k/p) = (\alpha, K \mid k) = 1$ (product formula for the norm residue symbol, [7; II, 26]).

The properties of this symbol give rise to local class field theory, i.e., to the description of Abelian extensions of fields $\mathbf{k}$ which are p-adically complete and are algebraic over some p-adic extension r_p of the rational field. If $\mathbf{k}$ is such a field and $\mathbf{K} \mid \mathbf{k}$ is Abelian, we define a local norm residue symbol $(\alpha, \mathbf{K} \mid \mathbf{k})$, for all α in $\mathbf{k}$, by choosing, according to Lemma 7, auxiliary fields K and k, algebraic over r, such that $K_P = \mathbf{K}$, $k_p = \mathbf{k}$, and the Galois groups are the same, and then

defining $(\alpha, \mathbf{K} \mid \mathbf{k})$ to be $(\alpha, K \mid k/p)$. Strictly, the latter symbol is an automorphism of $K \mid k$ only, while the former stands for the extension of this automorphism to $\mathbf{K} \mid \mathbf{k}$ by the method described in §1. It is easy to see that these auxiliary fields can be chosen in an infinite number of ways; until Theorem 6 is proved, we must therefore admit the possibility that the local norm residue symbol depends on the auxiliary fields used in constructing it. By following the definition of the norm residue symbol back to its genesis in the bracket symbol, one sees that, at a ramified p (the only ones causing any difficulty), it is defined according to the behavior of the extension field at all prime spots *except* p. So this definition of the local norm residue symbol is a very indirect one; a more direct definition, which avoids the auxiliary fields, can be given by means of simple algebras, but we will not discuss that here. (This other definition is described in [4] and [5; VII, §6].)

THEOREM 4. *If $K \mid k$ is Abelian, then $(\alpha_p, K \mid k/p)$ and $(\epsilon_p, K \mid k/p)$, respectively, take on as values exactly the elements of the decomposition group of p in $K \mid k$ and the inertia group of p in $K \mid k$.*

Proof. The decomposition field Z of p in $K \mid k$ is the largest subfield in which p splits completely; hence it follows from Theorem 3 that if $K \mid k$ is class field to $\mathfrak{h}$, then $Z \mid k$ is class field to $\mathfrak{a}_p\mathfrak{h}$; for $\mathfrak{a}_p\mathfrak{h}$ is the smallest idèle group including $\mathfrak{h}$ in which p splits completely. Thus, as in proof of Lemma 11, Z is the set of elements of K left invariant by the group of all automorphisms $(\mathfrak{a}_p, K \mid k)$; so this set of elements is the decomposition group. Similarly, the inertia field is class field to $\mathfrak{e}_p\mathfrak{h}$, so the set of $(\mathfrak{e}_p, K \mid k)$ is the inertia group.

COROLLARY 4.1. *If $K \mid k$ is Abelian and P is any prime divisor of p in K, then $\mathfrak{a}_p \cap \alpha N\mathfrak{A} = N\mathfrak{A}_P$; furthermore, $(\alpha_p, K \mid k/p) = 1$ if and only if α_p is norm of an element of K_P.*

Proof. Let α_p^* be the group of all elements of α_p for which the norm residue symbol is 1. Then by Theorem 4, $(\alpha_p : \alpha_p^*) = e_p f_p = (K_P : k_p)$. But $\alpha_p^* \supset N_P A_P$, and by Lemma 6, $(\alpha_p : N_P A_P) \leq (K_P : k_p)$. So $\alpha_p^* = N_P A_P$. Since $\mathfrak{a}_p \cap \alpha N\mathfrak{A}$ is the set of all $\mathfrak{a}_p$ whose p-component is in α_p^*, it is $N\mathfrak{A}_P$.

COROLLARY 4.2. *Suppose that $K \mid k$ is Abelian, S is a finite set of divisors, and a σ_p is chosen for every p such that*

(a) *σ_p is in the decomposition group of p in $K \mid k$,*
(b) *$\sigma_p = 1$ if p is not in S,*
(c) $\prod_p \sigma_p = 1$.

Then there exists an α in k such that $(\alpha, K \mid k/p) = \sigma_p$ at every p.

Proof. Define the k-idèle $\mathfrak{a}_p = (\alpha_p)$ by choosing α_p, for each p in S, to be an element of k_p for which $(\alpha_p, K \mid k/p) = \sigma_p$ (such elements exist by Theorem 4)

and choosing $\alpha_p = 1$ for p not in S. Then $(\mathfrak{a}, K \mid k) = 1$; so by the reciprocity law, $\mathfrak{a} = \alpha_0 N\mathfrak{A}_0$ for some α_0 and some $\mathfrak{A}_0$. $(\alpha_0, K \mid k) = (\alpha_p, K \mid k/p) = \sigma_p$ at every p.

THEOREM 5. *If* $\mathbf{k}$ *is* p*-adically complete and algebraic over* r_p, *and* $\mathbf{K} \mid \mathbf{k}$ *is Abelian, then* $(\alpha, \mathbf{K} \mid \mathbf{k})$ *is independent of the choice of auxiliary fields and gives an isomorphism between the Galois group of* $\mathbf{K} \mid \mathbf{k}$ *and the group of elements of* $\mathbf{k}$ *modulo the group of norms of elements of* $\mathbf{K}$.

Proof. It follows directly from Theorem 4 and the definition of the local norm residue symbol that there is such an isomorphism; only the uniqueness proof is non-trivial.

Let $K_1 \mid k_1$ and $K_2 \mid k_2$ be algebraic over r and contain prime divisors P_1, p_1 and P_2, p_2 such that $K_{iP_i} \cong \mathbf{K}$, $k_{ip_i} \cong \mathbf{k}$, and the Galois groups of $K_i \mid k_i$ are

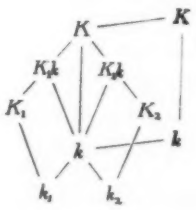


FIGURE 2

isomorphic to that of $\mathbf{K} \mid \mathbf{k}$. Since the K_i are both contained in $\mathbf{K}$, this field contains the product fields $K = K_1K_2$ and $k = k_1k_2$, and these fields possess prime divisors P and p, dividing P_i and p_i, respectively, such that $\mathbf{K} = K_P$ and $\mathbf{k} = k_p$ (see Fig. 2).

If α is any element of $\mathbf{k}$, let $\mathfrak{a}$, $\mathfrak{a}_1$, $\mathfrak{a}_2$ be idèles of k, k_1, k_2 with component α at p, p_1, p_2, respectively, and all other components 1. Since $k_p = k_{1p_1} = k_{2p_2}$, the p_i both split completely in $k \mid k_i$, so that $N_{k|k_1}\mathfrak{a} = \mathfrak{a}_1$ and $N_{k|k_2}\mathfrak{a} = \mathfrak{a}_2$.

From the definition of the local norm residue symbol it is clear that it suffices to prove that $(\mathfrak{a}_1, K_1 \mid k_1)$ and $(\mathfrak{a}_2, K_2 \mid k_2)$ give the same automorphism when extended to an automorphism of $\mathbf{K}$. Now $(\mathfrak{a}, K \mid k)$ and $(\mathfrak{a}, K_1k \mid k)$ both produce the same effect on elements of K_1; since $\mathbf{K}$ can be generated by adjoining elements of K_1 to $\mathbf{k}$, $(\mathfrak{a}, K \mid k)$ and $(\mathfrak{a}, K_1k \mid k)$ both extend to the same automorphism of $\mathbf{K}$. But $(\mathfrak{a}, K_1k \mid k) = (N_{k|k_1}\mathfrak{a}, K_1 \mid k_1) = (\mathfrak{a}_1, K_1 \mid k_1)$. Thus $(\mathfrak{a}_1, K_1 \mid k_1)$ gives the same automorphism of $\mathbf{K}$ as does $(\mathfrak{a}, K \mid k)$. By the same argument, we can show that this is also true for $(\mathfrak{a}_2, K_2 \mid k_2)$, and thus prove our theorem.

From Theorem 5 it follows at once that if $K_P \mid k_p$ is Abelian, then $(\alpha_p : N_P\mathrm{A}_P) = (K : k)$. So if we define $K_P \mid k_p$ to be a class field (to the group $N_P\mathrm{A}_P$) whenever these indices are equal, we get immediately that every Abelian extension of a p-adically complete field is a class field. The local norm residue symbol gives a

reciprocity law for such Abelian extensions; to complete local class field theory, we need only the existence theorem.

THEOREM 6. *If k_p is algebraic over r_p, and if γ_p is a subgroup of α_p such that $(\gamma_p : \alpha_p)$ is finite, then there is one and only one field K_P which is Abelian over k_p and is class field to γ_p (if we restrict ourselves to subfields of some fixed algebraically closed extension of k_p).*

COROLLARY 6.1. *If $K_P \mid k_p$ is any algebraic, complete field, $(\alpha_p : N_P\mathrm{A}_P)$ is equal to the degree over k_p of the largest subfield of K_P which is Abelian over k_p; every local class field is Abelian.*

Theorem 6 will follow from the much more inclusive Theorem 7. Corollary 6.1 is then proved just as in the large; if C_P is the Abelian class field to $N_P\mathrm{A}_P$, and A_0 is any element of A_P, then $(\mathrm{A}_0, C_PK_P \mid k_p) = (N_P\mathrm{A}_0, C_P \mid k_p) = 1$ since $N_P\mathrm{A}_0$ is by construction included among norms of elements of C_P. Hence $C_PK_P = K_P$ and $C_P \subset K_P$. The formula $(\mathrm{A}_P, C_PK_P \mid k_p) = (N_P\mathrm{A}_P, C_P \mid k_p)$ is true whenever $C_P \mid k_p$ is Abelian and $K_P \mid k_p$ algebraic; it follows at once from the similar formula for ideal elements.

6. Grunwald's theorem.

THEOREM 7. *If k is any algebraic number field, if S is a finite set of prime k-divisors, and if at each p of S a cyclic extension $\mathbf{K}(p)$, of finite degree over k_p, is given, then there exists a field $K \mid k$ such that*

(a) *$K \mid k$ is cyclic;*
(b) *if P is any prime K-divisor which divides p, and p is in S, then $K_P \cong \mathbf{K}(p)$;*
(c) *the degree of $K \mid k$ is the least common multiple of the degrees of the $\mathbf{K}(p) \mid k_p$.*

Furthermore, if for every p in S an element α_p is so chosen that $(\alpha_p, \mathbf{K}(p) \mid k_p)$ is a generator of the Galois group of $\mathbf{K}(p) \mid k_p$, then there is a field K which satisfies (a), (b), (c) *and has an automorphism σ such that*

(d) $(\alpha_p, K \mid k/p) = \sigma^{n/n(p)}$ *at each p in S, where $n = (K : k)$ and $n(p) = (\mathbf{K}(p) : k_p)$.*

To prove this, it suffices to prove

THEOREM 7′. *If to each p in S is assigned a group γ_p, such that α_p/γ_p is cyclic of finite order $n(p)$, and if $\mathfrak{h}_p$ is the group of all idèles with p-components in γ_p and other components* 1, *then there exists an idèle group $\mathfrak{h}$ such that*

(a′) *$\mathfrak{a}/\mathfrak{h}$ is cyclic, and $\mathfrak{h}$ is admissible;*
(b′) *if p is in S, $\mathfrak{a}_p \cap \mathfrak{h} = \mathfrak{h}_p$;*
(c′) *$(\mathfrak{a} : \mathfrak{h}) = n =$ least common multiple of the $n(p)$.*

Furthermore, if for each p in S an idèle $\mathfrak{c}_p$, which is a generator of $\mathfrak{a}_p$ (mod $\mathfrak{b}_p$), is chosen, there is an $\mathfrak{b}$ satisfying (a′), (b′), (c′) *and also*

(d′) *there is an idèle $\mathfrak{a}_0$ such that for each p in S $\mathfrak{c}_p \equiv \mathfrak{a}_0^{n/n(p)}$ (mod $\mathfrak{b}$).*

If the γ_p of Theorem 7′ are chosen as the groups to which, by Theorem 5, the fields $\mathbf{K}(p) \mid k_p$ are class fields, and if $K \mid k$ is the class field to $\mathfrak{b}$, then $K_P \mid k_p$ will be class field to γ_p; hence (b′) implies (b). The other conditions in Theorem 7 follow immediately from those in Theorem 7′.

We need the following elementary lemma on groups:

LEMMA 13. *Let g be a finite Abelian group, n an integer such that $g^n = (1)$, and h a subgroup of g such that*

$$g^\nu \cap h = h^\nu \tag{13}$$

for all $\nu \mid n$ (including $\nu = n$). Then g is the direct product of h and some subgroup h_1 of g.

Proof. By expressing g and h as direct products of subgroups of prime power order, it is easy to see that it suffices to prove the lemma for each such subgroup. Accordingly, we assume that the order of g is a power of the prime l.

Take h_1 to be any subgroup of minimal order such that $hh_1 = g$. To prove our lemma it suffices to show that $h_1 \cap h = (1)$. Suppose that this intersection contains an element $\eta \neq 1$. Let $\gamma_1, \cdots, \gamma_m$ be an independent basis for $\mathfrak{h}_1$ and let $\eta = \prod \gamma_\nu^{r_\nu}$, taking $r_i = 0$ unless the factor $\gamma_i^{r_i}$ is really $\neq 1$. Let l^ρ be the highest power of l (possibly l^0) which divides all the r_i, and assume that the r_i are so chosen that $r_1 = l^\rho$. Then by construction $\gamma_1^{r_1} \neq 1$.

But η is in $g^{r_1} \cap h$; so by (13), $\eta = \eta_1^{r_1}$, where η_1 is in h. Let

$$\gamma_1' = \gamma_1 \gamma_2^{r_2/r_1} \gamma_3^{r_3/r_1} \cdots \gamma_m^{r_m/r_1} \eta^{-1}$$

and consider the group h' generated by $\gamma_1', \gamma_2, \cdots, \gamma_m$. Clearly $g = h'h$; but since $(\gamma_1')^{r_1} = 1$ and $\gamma_1^{r_1} \neq 1$, the period of γ_1' is less than the period of γ_1. Hence h_1' is of smaller order than h_1. This is a contradiction.

Proof of Theorem 7′. Since the groups $\mathfrak{h}$ and $\mathfrak{h}_p$ are intersections of groups of prime power index in $\mathfrak{a}$ and $\mathfrak{a}_p$, respectively, it is easy to see that it suffices to prove the theorem for the case in which the $n(p)$ and their least common multiple n are powers of a single prime l. We will assume that this is the case.

The main tool in the proof is Lemma 12; after applying it, only manipulations with finite groups remain. By applying this lemma, then, there exists for each $\nu \mid n$ a finite set T_ν of prime k-divisors such that $\mathfrak{a}_{[S]} \cap \alpha\mathfrak{a}^\nu \mathfrak{b}^{S+T_\nu} = \mathfrak{a}_{[S]}^\nu$. Let T be the sum of the sets T_ν; then

$$\mathfrak{a}_{[S]} \cap \alpha\mathfrak{a}^\nu \mathfrak{b}^{S+T} = \mathfrak{a}_{[S]}^\nu \qquad \text{(for all } \nu \mid n\text{)}. \tag{14}$$

Let $\alpha\mathfrak{a}^n\mathfrak{b}^{S+T} = \mathfrak{h}_0$. Then (14) gives

$$\mathfrak{a}_{[S]} \cap \mathfrak{a}^\nu\mathfrak{h}_0 = \mathfrak{a}^\nu_{[S]} \qquad \text{(for all } \nu \mid n\text{)}, \tag{15}$$

$$\mathfrak{a}_{[S]} \cap \mathfrak{h}_0 = \mathfrak{a}^n_{[S]}\,. \tag{16}$$

We can now apply Lemma 13, with $g = \mathfrak{a}/\mathfrak{h}_0$ and $h = \mathfrak{a}_{[S]}\mathfrak{h}_0/\mathfrak{h}_0$; for these groups are finite, $g^n = (1)$, and, since $g^\nu \cap h$ is the group of all residue classes (mod $\mathfrak{h}$) generated by elements of $\mathfrak{a}_{[S]} \cap \mathfrak{a}^\nu\mathfrak{h}_0$, (15) shows that $g^\nu \cap h = h^\nu$ for all $\nu \mid n$. So $\mathfrak{a}/\mathfrak{h}_0$ is the direct product of $\mathfrak{a}_{[S]}\mathfrak{h}_0/\mathfrak{h}_0$ and another subgroup; that is, we can find an idèle-group $\mathfrak{h}_1$ such that $\mathfrak{a} \supset \mathfrak{h}_1 \supset \mathfrak{h}_0$ and $\mathfrak{a}/\mathfrak{h}_0$ is the direct product of $\mathfrak{a}_{[S]}\mathfrak{h}_0/\mathfrak{h}_0$ and $\mathfrak{h}_1/\mathfrak{h}_0$. Using this fact and (16), we see that

$$\mathfrak{a}/\mathfrak{h}_1 \cong \mathfrak{a}_{[S]}\mathfrak{h}_0/\mathfrak{h}_0 \cong \mathfrak{a}_{[S]}/(\mathfrak{a}_{[S]} \cap \mathfrak{h}_0) = \mathfrak{a}_{[S]}/\mathfrak{a}^n_{[S]}\,.$$

By use of this isomorphism we can reduce study of groups between $\mathfrak{a}$ and $\mathfrak{h}_1$ to study of groups between $\mathfrak{a}_{[S]}$ and $\mathfrak{a}^n_{[S]}$. First, let $\mathfrak{h}_{[S]}$ be the direct product of the

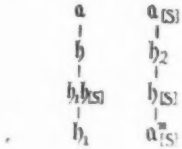


FIGURE 3

groups $\mathfrak{h}_p$ of Theorem 7′. Then since $\mathfrak{h}_{[S]} \supset \mathfrak{a}^n_{[S]}$, $\mathfrak{a}/\mathfrak{h}_{[S]}\mathfrak{h}_1 \cong \mathfrak{a}_{[S]}/\mathfrak{h}_{[S]}$. $\mathfrak{a}_{[S]}/\mathfrak{h}_{[S]}$ is the direct product of the groups $\mathfrak{a}_p/\mathfrak{h}_p$, and is thus generated by the elements $\mathfrak{c}_p$ of Theorem 7′, where p runs through S. To build $\mathfrak{h}$, we need only extend $\mathfrak{h}_{[S]}\mathfrak{h}_1$ to a suitable group whose factor group in $\mathfrak{a}$ is cyclic.

Since the period of each $\mathfrak{c}_p$ (mod $\mathfrak{h}_p$) is a power of l, there is one, say $\mathfrak{c}_{p_1}$, of the maximal period n. Take $\mathfrak{h}$ to be the group generated by the elements of $\mathfrak{h}_{[S]}\mathfrak{h}_1$ and the s - 1 elements

$$\mathfrak{c}_p\mathfrak{c}_{p_1}^{n/n(p)} \qquad (p \text{ in } S,\ p \neq p_1). \tag{17}$$

If $\mathfrak{h}_2$ is the subgroup of $\mathfrak{a}_{[S]}$ generated by the elements (17) and the elements of $\mathfrak{h}_{[S]}$, then $\mathfrak{a}/\mathfrak{h} \cong \mathfrak{a}_{[S]}\mathfrak{h}_1/\mathfrak{h}_2\mathfrak{h}_1 \cong \mathfrak{a}_{[S]}/\mathfrak{h}_2$. Hence $\mathfrak{a}/\mathfrak{h}$ is generated by $\mathfrak{c}_p$ and is cyclic of order n; $\mathfrak{c}_p \equiv \mathfrak{c}_{p_1}^{n/n(p)}$ (mod $\mathfrak{h}$), so (a′), (c′), and (d′) are verified (with $\mathfrak{a}_0 = \mathfrak{c}_{p_1}$). Finally, the elements (17) are so constructed that a product of powers of them is either in $\mathfrak{h}_{[S]}$ or differs from any element of $\mathfrak{h}_{[S]}$ for at least two p in S. Hence $\mathfrak{a}_p \cap \mathfrak{h}_2 = \mathfrak{a}_p \cap \mathfrak{h}_{[S]} = \mathfrak{h}_p$ and, since this implies $\mathfrak{a}_p \cap \mathfrak{h} = \mathfrak{h}_p$, (b′) is verified.

BIBLIOGRAPHY

1. C. CHEVALLEY, *Généralisation de la théorie du corps de classes pour les extensions infinies*, Journal de Mathématiques, series 9, vol. 15(1936), pp. 359–371.
2. C. CHEVALLEY, *La théorie du corps de classes*, Annals of Mathematics, vol. 41(1940), pp. 394–418.

3. C. Chevalley, *Sur la théorie du corps de classes dans les corps finis et les corps locaux*, Journal of the Faculty of Science, University of Tokyo, II, vol. 9(1933), pp. 366–476.
4. C. Chevalley, *La théorie du symbole de restes normiques*, Journal für die reine und angewandte Mathematik, vol. 169(1932), pp. 140–157.
5. M. Deuring, *Algebren*, Ergebnisse der Mathematik, vol. 4, no. 1, Berlin, Springer, 1935.
6. W. Grunwald, *Ein allgemeines Existenztheorem für algebraische Zahlkörper*, Journal für die reine und angewandte Mathematik, vol. 169(1932), pp. 103–107.
7. H. Hasse, *Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper*, parts I, Ia, and II, Jahresbericht des deutschen Mathematiker-Vereinigung, vol. 35(1926), pp. 1–55, vol. 36(1927), pp. 233–311, and Ergänzungsband VI, respectively.
8. K. Hensel, *Theorie der algebraischen Zahlen*, Leipzig and Berlin, Teubner, 1908.
9. F. K. Schmidt, *Zur Klassenkörpertheorie im Kleinen*, Journal für die reine und angewandte Mathematik, vol. 162(1930), pp. 155–168.
10. B. L. van der Waerden, *Elementarer Beweis eines Zahlentheoretischen Existenztheorems*, Journal für die reine und angewandte Mathematik, vol. 171(1934), pp. 1–3.
11. B. L. van der Waerden, *Moderne Algebra*, vol. 1, second edition, Berlin, Springer, 1937.
12. H. Weyl, *Algebraic theory of numbers*, Princeton University Press, 1940.
13. E. Witt, *Der Existenzsatz für abelsche Funktionenkörper*, Journal für die reine und angewandte Mathematik, vol. 173(1935), pp. 43–51.

Indiana University.

n-TO-ONE MAPPINGS OF LINEAR GRAPHS

By Paul W. Gilbert

1. **Introduction.** An n-to-1 continuous mapping is one for which every inverse image consists of exactly n points. Such mappings have been considered by O. G. Harrold [2], who showed that no 2-to-1 mapping can be defined on an arc. (In general, we shall use the term *mapping* to mean a continuous mapping.) J. H. Roberts [5] extended this result to a closed 2-cell and proved other theorems concerning 2-to-1 mappings defined over complete metric spaces. A paper by Roberts and Venable Martin [3] deals with such mappings of 2-dimensional manifolds. In a second paper [1] Harrold studied n-to-1 mappings on connected linear graphs.

Using the methods developed by Roberts, this paper considers first the question of defining a 2-to-1 mapping of any linear graph A. It is shown that unless the Euler characteristic $\chi(A)$ is even such a mapping cannot be defined on A. However, if $\chi(A)$ is odd, the following analogous question can be investigated. Does there exist a mapping of A which is 2-to-1 except that one inverse image consists of a single point? Γ is defined as the class of all mappings T defined over linear graphs, where T is either exactly 2-to-1 or else 2-to-1 except that one inverse image consists of a single point. In §3, it is shown that a mapping of class Γ can be defined on any linear graph which is a boundary curve and that any connected graph is the image of a boundary curve under some T belonging to Γ. In §4, the problem of the definition of n-to-1 mappings on a linear graph is considered. It is shown that if a mapping of class Γ can be defined on a linear graph A, then A admits an exactly n-to-1 mapping, for all $n \neq 2$.

2. **Two-to-one mappings.** Let T be an exactly 2-to-1 mapping defined over a linear graph A. (A linear graph is the sum of a finite number of arcs such that if a point p is common to two of the arcs, then p is an end point of each of them. Considering the end points as vertices and the arcs as 1-cells, we have a 1-dimensional complex.) The set of inverse images under T is an upper semi-continuous collection G of elements filling A, such that every element of G is a pair of points. For each point x in A, let $s(x)$ be the other point in the element. For any subset M of A, let $s(M)$ be the set of all points $s(x)$ for which x is in M. Let $f(x) = \rho(x, s(x))$, where ρ is the metric in A. Let K be the subset of A consisting of the points at which f is continuous. It follows from the upper semi-continuity of G that as x approaches a point q along an arc in K, $f(x)$ approaches

Received May 15, 1941; in revised form April 22, 1942; part of a doctoral dissertation presented June, 1940 at Duke University. The author is indebted to Professor Roberts, who suggested the problem and gave his advice and assistance during the preparation of this paper.

a limit which is either 0 or $f(q)$. The following results were obtained by Roberts [5; 257]:

(i) the functions $f(x)$ and $s(x)$ are continuous over exactly the same subset of A;
(ii) the set K is dense and open in A.

THEOREM 1. *Let C be an open connected subset of K, all the points of which are of second order in A. (By the order of a point we shall mean the Menger order.) Suppose there exists a point z in C such that $s(z)$ also belongs to C. Let $s(z) = w$ and let P be the arc zw in C. Then for every point x in $\overline{C} - P$, $s(x)$ must belong to P.*

If the theorem is false, then there exists a point q in $\overline{C} - P$ such that $s(q)$ does not belong to P. $\overline{C}$ may be an arc or a simple closed curve. In any case, we may assume without loss of generality that we have the order zwq on $\overline{C}$. Since $f(x)$ is positive and continuous over the closed set P, there exists an $\epsilon > 0$ such that $f(x) > \epsilon$ for x in P. Then the distance from z to w on the arc P is greater than ϵ. Let z_1 be the first point on this arc at a distance ϵ from z. Then $s(zz_1)$ is connected, by (i), and contains w. But q does not belong to $s(zz_1)$ since $s(q)$ is not in P. Also $s(zz_1)$ contains no point of zz_1 since $f(x) > \epsilon$ for x in zz_1. Hence $s(z_1)$ is between z_1 and q on the arc $Q = zwq$. If the distance from z_1 to w on the subarc z_1w of P is greater than ϵ, then let z_2 be the first point on the arc z_1w at a distance ϵ from z_1. Then, by the same argument, $s(z_2)$ is between z_2 and q on Q. After a finite number of steps, the point z_n is reached such that z_n is in P and is at a distance less than or equal to ϵ from w and such that $s(z_n)$ is between z_n and q. Then, as before, $s(w) = z$ must lie between w and q on Q. This is a contradiction.

COROLLARY 1. *If q is a point of order 2 in A and an end point of a component C of K and if $f(x) \to 0$ as $x \to q$ on C, then there exists a point p of C such that for any point z in the arc pq, $s(z)$ does not belong to pq.*

Pick the point p so close to q that the arc pq contains only second order points of K and does not contain $s(q)$. Suppose that z and $s(z) = w$ both belong to pq. Then by the theorem the point $s(q)$, in particular, must belong to the arc zw. This is a contradiction.

COROLLARY 2. *Let M be any arc of A, all of whose points are of order 2 in A. If there exists a component C of K contained in M, with end points p and q, such that $f(x) \to 0$ as $x \to p$ and as $x \to q$ on C, then for every point x in $M - \overline{C}$, $s(x)$ belongs to C.*

Suppose there exists a point r in $M - \overline{C}$ such that $s(r)$ is not in C. We may assume that we have the order pqr on M. Now $f(x) \to 0$ as $x \to q$ on C; hence by Corollary 1 there are points of $s(C)$ on the open arc segment qr. By (i), $s(C)$ is connected. Now $C \cdot s(C) = 0$. For, otherwise, the points z and $w = s(z)$ would both belong to C and then, by the theorem, for every point x of $\overline{C} - zw$, $s(x)$ would be in the arc zw, which is impossible. Therefore, since it does not

intersect r, $s(C)$ is a subset of qr. But as $x \to p$ in C, $s(x) \to p$. This is a contradiction.

THEOREM 2. *If q is a point of order 2 in A and an end point of a component C of K, then $f(x) \to 0$ as $x \to q$ on C.*

Suppose the contrary. Then $f(x) \to f(q)$ as $x \to q$ on C. Let U be an open arc segment of A containing q and consisting entirely of second order points of A. By a theorem of Roberts [5; 258], there exists in U an arc p_1q_1 and an open arc segment W such that

(1) p_1q_1 is in W;
(2) q_1 does not belong to K, but $p_1q_1 - q_1$ is in K;
(3) $f(x) \to f(q_1) = 4\epsilon$ as $x \to q_1$ on p_1q_1 ;
(4) if x is in W, then either $f(x) < \epsilon$ or $f(x) > 3\epsilon$;
(5) if p_2q_2 is any arc satisfying (1) and (2) and if $f(p_2) < \epsilon$, then $f(x) \to 0$ as $x \to q_2$ on the arc p_2q_2 .

Now q_1 is a limit point of points of discontinuity of f. For, in the contrary case, there exists a component C_1 of K having q_1 as an end point such that $f(x) \to 0$ as $x \to q_1$ on C_1 . But then, by Corollary 1, as $x \to q_1$ on C_1 , $s(x) \to q_1$ but does not lie in C_1 . This is a contradiction since $f(x) \to f(q_1)$ as $x \to q_1$ on p_1q_1 . Now since q_1 is a point of discontinuity, there exists a sequence of points $\{x_n\}$ such that $x_n \to q$ and $f(x_n) \to 0$ as $n \to \infty$. Choose j so large that $f(x_j) < \epsilon$, x_j is in W, and x_j lies between two discontinuities which are in W. By the upper semi-continuity there exists an open subsegment W_1 of W containing x_j , such that $f(x) < \epsilon$ for all x in W_1 . Since K is dense on A, there exists a point p_2 in $K \cdot W_1$. Let C_2 be the component of K containing p_2 , and denote its end points by y and z. Then $f(x) < \epsilon$ for x in C_2 , and $f(x) \to 0$ as $x \to y$ and as $x \to z$ on C_2 . But since $f(q_1) = 4\epsilon$ and $f(x) < \epsilon$ on C_2 , $s(q_1)$ does not belong to C_2 . This contradicts Corollary 2.

THEOREM 3. *Let M be an open connected subset of A consisting of points of order 2 in A. Then in M there are at most two points of discontinuity of the function $f(x)$.*

In the contrary case, let C be a component of K in M, with end points p and q in M. Then by Theorem 2, as $x \to p$ and as $x \to q$ on C, $f(x) \to 0$. Now there exists another point r of M which does not belong to K. Hence there exists a sequence of points $\{x_n\}$ such that x_n is in M, $x_n \to r$ and $f(x_n) \to 0$ as $n \to \infty$. Choose j so large that $f(x_j) < \rho(x_j , \overline{C})$. Then $s(x_j)$ does not belong to C. This contradicts Corollary 2.

DEFINITION. Given a mapping h defined over a set M, a subset of M will be called *integral* with respect to h, provided it is such that, if it contains one point of $h^{-1}(y)$, it contains every point of $h^{-1}(y)$ [4].

DEFINITION. Let H be the smallest subset of the linear graph A which is integral with respect to T and contains the vertices of A and the points of $A - K$.

By Theorem 3, the set H is a finite point set. Taking the points of H as vertices, the linear graph A gives rise to a 1-dimensional geometric complex, which will be denoted by X. The Euler characteristic $\chi(A)$ for the linear graph can be defined with respect to this complex as $\chi(A) = a_0 - a_1$, where a_0 is the number of vertices and a_1 the number of 1-cells in the complex.

THEOREM 4. *A necessary condition for the existence of a 2-to-1 mapping T of the linear graph A is that the Euler characteristic $\chi(A)$ be even.*

Let X be the complex determined by the point set H. For any point z in an open 1-cell P of X, $s(z) = w$ cannot belong to P. For if w were in P, then the subarc zw of P would be composed of points of K of order 2. But then, by Theorem 1, for every point x in $\overline{P} - zw$, $s(x)$ would be in zw. This is a contradiction since, in particular, the end points of P belong to the integral set H. Therefore, the function $s(x)$ is topological on P. Hence $s(P)$ is an open arc and contains no points of H. Since the end points of $s(P)$ must be in the set H, it follows that $s(P)$ is also an open 1-cell of X. Thus the mapping T determines a pairing of the 1-cells and vertices of X, and, therefore, the Euler characteristic is even.

Harrold [1; Theorems 4.3 and 4.4] has proved that if T is a 2-to-1 mapping of a graph A, $T(A) = B$, and $T^{-1}(y) = p + q$ for y in B, then B is a linear graph and

$$2 \cdot o(y) = o(p) + o(q), \tag{1}$$

where $o(x)$ denotes the order of the point x. (Harrold's theorems are stated for connected linear graphs, but their truth for any linear graph can readily be established.) He remarks [1; footnote 11] that formula (1) implies that $\chi(A) = 2\chi(B)$. It follows from (1) also that the orders of x and $s(x)$ are either both even or both odd. Harrold's theorem [2] that no 2-to-1 mapping can be defined on an arc is an immediate consequence of Theorem 4.

THEOREM 5. *If there exists a mapping T^* defined on A which is 2-to-1, except that one inverse image consists of a single point, then $\chi(A)$ is odd.*

Let z be the single unpaired point of A. For the mapping T^*, define $s^*(x)$ and $f^*(x)$ as s and f were defined for the exactly 2-to-1 mapping T, taking $s^*(z) = z$. As $x \to z$, $f^*(x) \to 0$, since otherwise there would exist a point $s^*(z) \neq z$. (Note that z must be of even order.) Define the point set H^* relative to T^* as H was defined relative to T, but let z belong to H^*. Then the points of H^* determine a complex X^*. Now the mapping T^* is topological on every open 1-cell P of X^*, and hence $T^*(P)$ is an open arc with end points in the finite set $T^*(H^*)$ (these end points may be either distinct or coincident points). If P and Q are two open 1-cells of X^*, either $T^*(P) = T^*(Q)$ or $T^*(P) \cdot T^*(Q) = 0$. For suppose y is in $T^*(P) \cdot T^*(Q)$. Write $T^{*-1}(y) = x + s^*(x)$, where x is in P, $s^*(x)$ is in Q. Then it follows that $s^*(P) = Q$. Therefore, X^* contains an even number of 1-cells. But the number of vertices is odd, and hence $\chi(A)$ is odd.

Corresponding to the problem of defining a 2-to-1 mapping on a graph with even characteristic, there is the problem of defining on a graph with odd characteristic a mapping which is 2-to-1 with one exception. The term 2-to-1 *with one exception* will be used to indicate a mapping which is 2-to-1 except that one inverse image consists of a single point.

DEFINITION. Let Γ be the class of all mappings, defined on linear graphs, which are either exactly 2-to-1 or else are 2-to-1 with one exception.

DEFINITION. Let $E(A) = -\chi(A)$.

In the next sections T, which has been used in this section to denote an exactly 2-to-1 mapping, will indicate any mapping in the class Γ. The functions $s(x)$, $f(x)$ and the sets K, H are defined as for the exactly 2-to-1 mapping. But if there exists an unpaired point z of A, then we take $s(z) = z$ and let H include the point z.

The examples that follow are linear graphs on which no mapping of class Γ can be defined. The proofs are omitted.

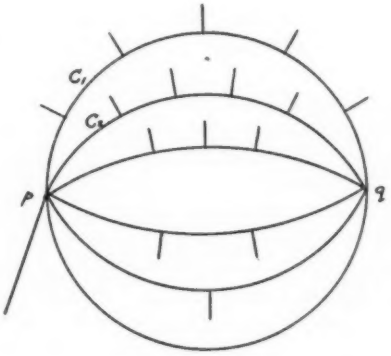


FIGURE 1

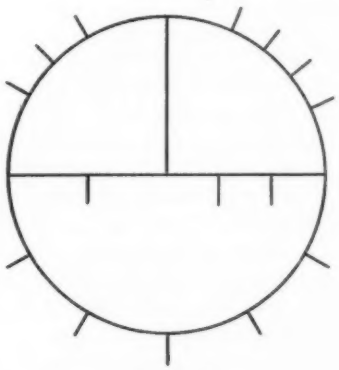


FIGURE 2

Example 1. Let A_1 be the graph in Fig. 1. It consists of 37 1-cells and 33 vertices, and, therefore, $E(A_1) = 4$. But no mapping of class Γ can be defined on A_1.

Example 2. Let A_2 be the graph $A_1 - C_1$ in Fig. 1 (C_1 is the indicated component of $A_1 - (p + q)$). Then $E(A_2) = 3$, but A_2 does not admit a mapping of class Γ. However, the graph $A_2 - C_2$ does admit such a mapping.

Example 3. In Example 2, $E(A_2) = 3$. But there exist connected graphs, for which the value of E is less than 3, on which no mapping of class Γ can be defined. An example of such a graph is given in Fig. 2. The value of E for this graph is 2. Note also that this example contains no point of order greater than 3.

It will be shown in §3 that if A is a connected linear graph and $E(A) < 2$, then a mapping of class Γ can always be defined on A.

3. **Boundary curves.** (A boundary curve is a locally connected continuum such that each of its true cyclic elements is a simple closed curve.)

THEOREM 6. *If a linear graph A is a boundary curve, then a mapping T belonging to the class Γ can be defined on A.*

The mapping is determined by defining over A a transformation s which has the following properties:

(1) for every point x in A, $s(x)$ is a point in A;

(2) s is continuous except at a finite number of points of A;

(3) if $s(x) = y$, then $s(y) = x$, except that one point z may be such that $s(z) = z$;

(4) the collection G filling A, each of whose elements is the set of points x, $s(x)$, is upper semi-continuous. Then T is defined as that mapping of A into G such that each point of A is carried into the element of G to which it belongs. In the proof we distinguish three cases.

Case 1. $E(A) = -1$. Then the graph is acyclic. The proof proceeds by induction on the number of 1-cells of the graph. If A consists of one 1-cell, with end points p, q, then pick any other point r in A. Let the function s map pr topologically into rq, with $s(p) = q$ and $s(r) = r$. If for x in rq we let $s(x) = s^{-1}(x)$, then s is completely defined over A. Suppose now that A consists of n 1-cells, where $n \geq 2$ (assume that only points of order different from 2 are vertices of A). Then there exists a pair of 1-cells pq and pr of A such that q and r are points of first order in A. Let s be a topological mapping of pq into pr such that $s(q) = r$ and $s(p) = p$. For x in pr, let $s(x) = s^{-1}(x)$. Let $A_1 = A - (pq + pr) + p$. By the induction, the s-function can be defined on A_1. Hence s is completely defined on A, and the induction is complete.

Case 2. $E(A) = 0$. Then A contains just one simple closed curve J. In this case it will be shown that, given any point z in J which is of second order in A, there exists a mapping T of class Γ defined on A such that as $x \to z$ on any arc of A, $f(x) \to 0$.

If there exists a pair of 1-cells pq and pr of A such that q and r are points of first order in A, then s can be defined on these two arcs as it was in Case 1, leaving the point p unpaired. Repeat this process on the subgraph A_1 of A on which s is still undefined, if A_1 contains such a pair of arcs. (A subgraph A_1 of A is a graph containing a subset of the 1-cells of A and the vertices of A which are on those 1-cells.) After a finite number of steps, we obtain in this way a subgraph A^* of A such that $A^* = J + (p_1q_1 + p_2q_2 + \cdots + p_mq_m)$, where the arcs $\{p_iq_i\}$ have no points in common and each has only the point p_i in common with J. (A^* may equal J, in which case $m = 0$.) It remains to define s on A^*.

We distinguish two cases. If m is even or 0, there exists a point z^* of J, which is of second order in A^* and which is such that the two components C_1 and C_2 of $A^* - (z + z^*)$ contain the same number of points of third order. Then C_1 and C_2 are homeomorphic. Let s be a topological mapping of $\overline{C}_1$ into $\overline{C}_2$ such

that $s(z) = z$, $s(z^*) = z^*$. For x in C_2, let $s(x) = s^{-1}(x)$. The definition of s on A^* is completed in this case by taking $s(z) = z^*$. If, on the other hand, m is odd, pick the vertex p_i such that the two components C_1 and C_2 of $A^* - (z + p_iq_i)$ contain the same number of points of third order. Since C_1 and C_2 are homeomorphic, s may be taken as a topological mapping of $\overline{C}_1$ into $\overline{C}_2$, with $s(z) = z$ and $s(p_i) = p_i$. Pick any interior point r_i of the arc p_iq_i, and let s map p_ir_i topologically into r_iq_i, with $s(p_i) = q_i$. Complete the definition of s by taking $s(z) = r_i$. (Note that $f(x) \to 0$ as $x \to z$ and as $x \to s(z)$ on each arc of A. A point such as z will be called a *free point* relative to s, since it may be paired with any other such point, or left unpaired.)

Case 3. $E(A) > 0$. Let J be any simple closed curve in A. Let C be the largest connected subgraph of A which contains J and which contains at most one point from any other simple closed curve in A. Then $E(C) = 0$. Hence, by Case 2, the s-function can be defined on C. Let $A_1 = \overline{A - C}$. The subgraph C contains just one point in each of a finite set of simple closed curves $J_1, J_2, \cdots, J_k$. Denote these points by $x_1, x_2, \cdots, x_k$. Then let C_i be the largest connected subgraph of A_1 which contains J_i and which contains at most one point from any other simple closed curve in A_1 ($j = 1, 2, \cdots, k$). Since the maximal cyclic elements of A are simple closed curves, then the sets C_1, $C_2, \cdots, C_k$ have no points in common. Moreover, the set $D = C_1 + C_2 + \cdots + C_k$ contains at most one point from any simple closed curve of A_1 which is not contained entirely in D. From the definition of C, the point x_j must be of second order in C_j. Moreover, $E(C_j) = 0$. Hence, by Case 2, the function s can be defined over C_j so that x_j and $s(x_j)$ are free points relative to s. Now x_j has already been paired with a point on C, and hence we shall leave $s(x_j)$ unpaired for the present. Thus s is defined over each set C_j in D. If the graph $A_2 = \overline{A_1 - D}$ is not vacuous, then repeat the process on A_2. After a finite number of steps, the graph A is exhausted. Then the s-function has been defined everywhere on A, except for a single unpaired point on each simple closed curve of A (except J). Since these points are all free points relative to s, they may be paired up arbitrarily. Note that if $E(A)$ is even, then there are an odd number of simple closed curves in A, and hence the mapping is exactly 2-to-1. If $E(A)$ is odd, a single point is left unpaired. This completes the proof of the theorem.

Note that the problem of the definition of T belonging to Γ on any connected graph A for which $E(A) = -1$ or 0 is solved by Cases 1 and 2 of this theorem. In fact, we have the following:

THEOREM 7. *If A is a connected linear graph and $E(A)$ is less than* 2, *then a mapping T of class Γ can be defined on A.*

Since A is connected, $E(A)$ must take on one of the values -1, 0, or 1. Then either (1) A is a boundary curve, or (2) A contains just one true cyclic element, which is a θ-curve. In (1), A admits a mapping of class Γ, by the preceding

theorem. In (2), denote by J the θ-curve in A. By the method used in Case 2 of Theorem 6, we can reduce the problem to the definition of the s-function on a subgraph A^* of A such that $A^* = J + (p_1q_1 + p_2q_2 + \cdots + p_mq_m)$, where the arcs $\{p_iq_i\}$ have no points in common and each has only the point p_i in common with J. Denote the three arcs of J by pxq, pyq, pzq. Suppose first that all the points of these arcs, except p and q, are of second order in A^*. Then let s be a topological mapping of pxq into pyq, with $s(p) = p$ and $s(q) = q$. The subgraph of A^*, on which s has not been defined, is acyclic, and hence s may be defined on it by Case 1 of Theorem 6. Suppose, on the other hand, that one of the three open arcs in J, say pxq, has points of third order in A^*. Let r be the last point of third order on this open arc. Let C be the component of $A^* - (p + r)$ which contains the point q. Then $E(\overline{C}) = 0$, and p is of order 2 in $\overline{C}$. Hence, by Case 2 of Theorem 6, s may be defined on $\overline{C}$ so that p is a free point relative to s. The graph $D = A^* - C$ is acyclic, and r is of order 2 in D. Hence, by an extension of Case 1 of Theorem 6, s can be defined on D so that r is a free point relative to s. This defines s completely on A^*.

If the graph A were not connected, this theorem would not be true. If fact, let A consist of three components, two of which are simple arcs and the third of which is the graph A_2 given in Example 2 of §2. Then $E(A) = 1$, but it can be shown that A does not admit a mapping of class Γ.

THEOREM 8. *Let M be the class of all linear graphs A which are boundary curves. Let N be the class of all image spaces $T(A)$, where A belongs to M and T belongs to Γ. Then N is the class of all connected linear graphs.*

By Theorem 6, if A is in M there exists a mapping T of class Γ defined on A. By an application of the procedure used in §2 on exactly 2-to-1 mappings, it can be shown that $T(A)$ is a linear graph. From the continuity of T, since a boundary curve is connected, then $T(A)$ is connected. It remains only to prove that, if B is any connected linear graph, then there exists a graph A belonging to M and a mapping T belonging to Γ such that $T(A) = B$. It will be enough to define T so that $T(A)$ is homeomorphic to B, not necessarily equal to B. For if we denote the homeomorphism by h, then $hT(A) = B$, and the mapping hT will belong to Γ if T does.

As in the proof of Theorem 6, the s-function will first be defined on A. This function gives rise to an upper semi-continuous collection G filling A. Then T will be taken as that mapping of A into G which carries a point x of A into the element of G to which it belongs. The proof proceeds by induction on the number of 1-cells in B.

Suppose first that B consists of a single 1-cell, i.e., B is an arc. Then let A be an arc pq, and pick any interior point r of this arc. Let s be a topological mapping of pr into rq such that $s(p) = q$ and $s(r) = r$. For x in rq, let $s(x) = s^{-1}(x)$. Then $T(A)$ is an arc, and hence is homeomorphic to B.

Now let B be any connected linear graph composed of n 1-cells ($n > 1$). Two cases will be distinguished.

Case 1. *B is acyclic.* Then there exists a 1-cell P of B such that $B_1 = \overline{B - P}$ is connected. Hence P has only one end point y in common with B_1. By the induction there exists A_1 belonging to M and T_1 belonging to Γ such that $T_1(A_1) = B_1$. Let s_1 be the s-function corresponding to the mapping T_1. Determine z in A_1 such that $T_1(z) = y$. The graph A is formed from A_1 by adding independent arcs zp and zq, which have only the end point z in common with A_1. Let s map zp topologically into zq, with $s(p) = q$ and $s(z) = z$. For x in A_1, let $s(x) = s_1(x)$. Then the mapping T corresponding to the s-function is such that $T(A)$ is homeomorphic to B.

Case 2. *B contains a simple closed curve J.* Let P be a 1-cell of B contained in J. Then $B_1 = \overline{B - P}$ is connected and P has only its two end points y_1 and y_2 in common with B_1. By the induction there exists A_1 in M and T_1 in Γ such that $T_1(A_1) = B_1$. Determine x_1 and x_2 in A_1 such that $T_1(x_1) = y_1$ and $T_1(x_2) = y_2$. Then A is formed from A_1 by adding the mutually exclusive simple closed curves $x_1ap_1bx_1$ and $x_2cp_2dx_2$, which have only the points x_1 and x_2 in common with A_1. Let s be a topological mapping of x_1ap_1 into x_1bp_1 such that $s(x_1) = x_1$ and $s(p_1) = p_1$. Define s similarly on the other simple closed curve and set $s(p_1) = p_2$. For x in A_1, let $s(x) = s_1(x)$. The corresponding mapping T is such that $T(A)$ is homeomorphic to B. Thus the induction is complete.

Note that the mapping T defined in the course of the induction is exactly 2-to-1, or 2-to-1 with one exception, according as T_1 is. But the mapping, in case B is a single 1-cell, is 2-to-1 with one exception. It follows that, given any connected linear graph B, there exists A belonging to M and a mapping T which is 2-to-1 with one exception such that $T(A) = B$. This last statement is not true if we take T to be an exactly 2-to-1 mapping. But it can be shown that if the connected graph B contains a simple closed curve, then there does exist A in M and an exactly 2-to-1 mapping T such that $T(A) = B$.

4. **n-to-1 mappings.** We consider now the more general case in which a mapping W of a linear graph A is such that every inverse image consists of exactly n points, where n is an arbitrary positive integer. We define first three particular transformations of an arc A into itself. In these three examples, n will denote an odd integer.

Type 1. Let the end points of A be a_0 and a_n. Pick $n - 1$ distinct interior points of A, and label them so that we have the order $a_0, a_1, a_2, \cdots, a_n$ on A. Let W_1 be a transformation of A into itself having the following properties:

(1) $W_1(a_i) = a_0$ if i is even; $W_1(a_i) = a_n$ if i is odd;

(2) for each i, W_1 is a topological mapping of the arc a_ia_{i+1}, carrying this arc into A. Then each inverse image under W_1 consists of exactly n points, except that $W_1^{-1}(a_0)$ and $W_1^{-1}(a_n)$ consist of $(n + 1)/2$ points each.

Type 2. Denote the end points of A by x_1 and x. Pick in the interior of A a sequence of distinct points $x_2, x_3, \cdots$ which are in the indicated order on A and

have x as a sequential limit point. We define a transformation W_2 of A into itself having the following properties:

(1) $W_2(x) = x$, and for each i, $W_2(x_i) = x_i$;

(2) for each i, W_2 is a transformation of Type 1 on the arc x_ix_{i+1}, carrying this arc into itself. Then each inverse image under W_2 consists of exactly n points, except those containing x_1 and x. $W_2^{-1}(x)$ consists of a single point, and $W_2^{-1}(x_1)$ contains $(n + 1)/2$ points.

Type 3. Let A be the arc zx_1x. We define a transformation W_3 of A into itself as follows. On the arc x_1x, W_3 is a transformation of Type 2, carrying x_1x into itself, with $W_3^{-1}(x) = x$. Similarly on x_1z, W_3 is of Type 2, with $W_3^{-1}(z) = z$. Then each inverse image under W_3 consists of n points, except that $W_3^{-1}(x)$ and $W_3^{-1}(z)$ each consist of a single point.

These mappings will be referred to for convenience as n-to-1 mappings of Types 1, 2, and 3 although they are not exactly n-to-1.

THEOREM 9. *If n is odd, then every linear graph A admits an n-to-1 mapping.*

To the vertices of A add enough second order points so that the number of points in the resulting finite point set V is exactly divisible by n. We define a continuous transformation W_1 on A as follows. On each 1-cell pq of the complex determined by the set V, let W_1 be an n-to-1 mapping of Type 3 carrying pq into itself, with $W_1(p) = p$. Then $W_1(A) = A$, and every inverse image under W_1 consists of n points, except that the points of V are left unpaired. Group these points into mutually exclusive sets V_1 , V_2 , $\cdots$, V_i , consisting of n points each. Let W_2 be the mapping of A obtained by identifying all the points in each of these sets V_i . Then the mapping $W = W_2W_1$ is an exactly n-to-1 mapping of A. The image space in this case is a linear graph.

THEOREM 10. *If a linear graph A admits a mapping T of class* 1, *then there exists an n-to-1 mapping of A, for every $n \neq 2$.*

By the preceding theorem, we need consider only the case where n is even. An n-to-1 mapping W_2 will be defined over A. Let L be the smallest integral subset of A satisfying the following conditions:

(1) L contains the point set H which corresponds to the mapping T (see §2);

(2) if A contains an unpaired point z (i.e., $s(z) = z$), then the component of $A - L$ which contains z contains no other point of H;

(3) the number of pairs of points p, $s(p)$ in L is even (if A contains an unpaired point z, then z will be considered as one of these pairs of points in L).

Then L is a finite point set. If we take the points of L as vertices, A gives rise to a complex X. The mapping T determines a pairing of the 1-cells of X.

Let P be an open 1-cell of X, with end points p and q, and let $Q = s(P)$. One end point of Q is either p or $s(p)$, the other either q or $s(q)$. On $\overline{P}$ define W as an

$(n-3)$-to-1 mapping of Type 3 carrying $\overline{P}$ into itself, with $W(p) = p$, $W(q) = q$. On $\overline{Q}$ define W as a 3-to-1 mapping of Type 3 (only in this case a mapping of $\overline{Q}$ into $\overline{P}$) such that if p or q belongs to $\overline{Q}$ then they are each self-corresponding, and if $s(p)$ or $s(q)$ belongs to $\overline{Q}$, then $W(s(p)) = p$, $W(s(q)) = q$. Define W in this way for every pair of 1-cells which correspond under s. Finally, if p and $s(p)$ are free points relative to s (see end of Case 2 of Theorem 6), make the added condition that W identifies these two points. Then all the inverse images under W consist of n points, except for those containing points of L. For p in L, the points p and $s(p)$ map into the same point under W.

We define now a new mapping W_1 of A by amending the definition of W as follows. Consider a pair of distinct points p, $s(p)$ of L. By the definition of L there exists an open 1-cell P, of which W gives an $(n-3)$-to-1 mapping of Type 3, such that one of the pair p, $s(p)$ (say, for definiteness, p) is one end point of P and such that the other end point of P is not the unpaired point z of A (in case z exists). On $\overline{P}$ define W_1 as an $(n-3)$-to-1 mapping of Type 2 carrying $\overline{P}$ into itself, where the inverse image containing p consists of $(n-2)/2$ points. Repeat this process for every pair of distinct points p, $s(p)$ of L (if in the course of this procedure we come to a 1-cell $\overline{P}$ for which a mapping of Type 2 has already been introduced, then we let W_1 be a mapping of Type 1 of this 1-cell into itself). Suppose finally that there exists a point z which is unpaired in the mapping T. Let R be any open 1-cell having z as one end point and let $S = s(R)$. On $\overline{R}$ let W_1 be an $(n-1)$-to-1 mapping of Type 2 carrying $\overline{R}$ into itself, such that the inverse image containing z consists of $n/2$ points; on $\overline{S}$ let W_1 be a 1-to-1 mapping of $\overline{S}$ into $\overline{R}$, with $W_1(z) = z$. Everywhere else on A let W_1 be identical with W. Then we have defined a mapping W_1 such that each pair of points p, $s(p)$ of L belongs to an inverse image consisting of $n/2$ points, while all other inverse images consist of n points each.

Denote by $L_1, L_2, \cdots, L_j$ the inverse images containing the points of L. By the definition of L, there are an even number of these sets L_i. Each consists of $n/2$ points. We define a new mapping W_2 of A which is identical with W_1, except that it identifies the sets L_i in pairs. Then W_2 is exactly n-to-1.

Note that, under the mapping W_2 as described in this theorem, $W_2(A)$ is a linear graph. The existence of a mapping of class Γ on A is not a necessary condition for the existence of an n-to-1 mapping of A for all $n \neq 2$. In fact, for every n, an n-to-1 mapping can be defined on the graph A_2 given in Example 2 of §2.

For any value of n other than 2, no example has yet been given of a graph which does not admit an n-to-1 mapping. Also the fundamental difference, if any, between even-to-1 and odd-to-1 mappings has not yet been investigated. (By an even-to-1 mapping is meant an n-to-1 mapping where n is an even integer.)

Bibliography

1. O. G. Harrold, Jr., *Exactly* $(k,1)$ *transformations on connected linear graphs*, American Journal of Mathematics, vol. 62(1940), pp. 823–834.

2. O. G. Harrold, Jr., *The non-existence of a certain type of continuous transformation*, Duke Mathematical Journal, vol. 5(1939), pp. 789–793.
3. Venable Martin and J. H. Roberts, *Two-to-one transformations on 2-manifolds*, Transactions of the American Mathematical Society, vol. 49(1941), pp. 1–17.
4. J. H. Roberts, *Concerning collections of continua not all bounded*, American Journal of Mathematics, vol. 52(1930), p. 554.
5. J. H. Roberts, *Two-to-one transformations*, Duke Mathematical Journal, vol. 6(1940), pp. 256–262.

Texas Technological College.

MONOTONE TRANSFORMATIONS

By A. D. Wallace

In this paper we initiate the study of transformations which are monotone relative to a collection of sets. These mappings include, by proper specialization of the family of sets, both monotone and non-alternating transformations. Since we are concerned with non-metric spaces it is necessary to extend to such spaces those results of the classical theory of continua needed in our treatment. Where available proofs were adequate we have made reference to them. The first section contains general results on continua and, in particular, a proof, for non-separable spaces, of the theorem that every continuum contains at least two non-cutpoints. In the second section we give a cyclic element theory and show that if p is neither an endpoint nor a cutpoint then there is a point conjugate to p. Certain essential departures from the situation in separable spaces are noted. In the last section an extension of the result (due to Schweigert) that A-sets are invariant under non-alternating transformations is given.

1. By a topological space we mean a set S together with a class of closed sets satisfying the conditions CS of Lefschetz [7; I]. In general we adhere to the terminology of Lefschetz and Alexandroff-Hopf [1]. However, we use "compact" in the sense of "bi-compact" and in a normal space it is not necessarily assumed that a single point is a closed set.

A family of sets will be termed an *M-collection* if each set is closed and if, with each pair of sets, the collection contains their intersection. It is clear that an aggregate of closed sets is an M-collection if and only if, with each finite family of sets, it contains their intersection. A property B of closed sets will be termed *inductive* provided that, if each element of an M-collection has property B, then the intersection of all the sets has property B. Equivalently we also speak of a *collection G* as being *inductive* if each M-collection in G has its intersection in G. Manifestly a theorem about inductive collections implies a theorem about inductive properties and conversely, so that the two concepts may be used correlatively.

By a partially ordered system we mean a class P and a binary relation R between elements of P such that (i) pRp for each p in P, (ii) pRq and qRp imply $p = q$, (iii) pRq and qRr imply pRr. A subset P_0 of P will be called an ordered system if for each pair of elements p and q in P_0 we have either pRq or qRp. It is known [4; 140] that any ordered subsystem P_0 of P is contained in a *maximal* ordered subsystem of P. This result is of fundamental importance in non-separable spaces.

(1.1) *Any non-null inductive collection G of S contains a minimal element.*

Received December 5, 1941; revised May 18, 1942; presented to the American Mathematical Society December, 1939 and September, 1941.

Proof. Let P be the system of all M-collections of G, partially ordered by inclusion. It is readily seen that P contains a non-void ordered system and hence P contains a maximal ordered sub-system P_1. Let G_0 be the union of all the collections in P_1. Then it is manifest that G_0 is a maximal M-collection of G. The intersection of all the sets in G_0 is the desired minimal set.

This result may also be stated as

(1.2) *If S has an inductive property B, then S contains a minimal closed set having property B.*

It is clear that (1.2) is an extension of the well-known reduction theorem of Brouwer. For related results see [10], [15] and, in particular, [12; 84].

We see at once that

(1.3) *In order that a topological space be compact it is necessary and sufficient that the property of being non-null be inductive.*

Let G be a non-empty collection of closed sets of S. We denote by $M(G)$ the family of all closed sets Y in S such that XY is an element of G for each X in G.

(1.4) *For any non-void collection G of closed sets we have:*

(1.4.1) *$M(G)$ is an M-collection;*
(1.4.2) *S is an element of $M(G)$;*
(1.4.3) *S is in G if and only if $M(G) \subset G$;*
(1.4.4) *$M(M(G)) = M(G)$;*
(1.4.5) *if G is an inductive collection, then so is $M(G)$.*

Proof. Let $G = [X]$. If Y and Z are in $M(G)$ we have XY in G for X in G. Hence $Z(XY)$ is in G and so $ZY \in M(G)$. It is clear that (1.4.2) holds. Also (1.4.3) is easily proved and implies that $M(M(G)) \subset M(G)$. Now it is clear that for any M-collection G' we must have $G' \subset M(G')$. Since $M(G)$ is an M-collection we thus have $M(G) \subset M(M(G))$, and this completes the proof of (1.4.4). To prove that (1.4.5) holds, let $H = [Y]$ be an M-collection of elements of $M(G)$ and let Z be the intersection of the elements of H. We have to show that $ZX \in G$. Now ZX is the intersection of the elements of the family $[YX]$. Since G is an inductive collection and $XY \in G$ we must prove that $[XY]$ is an M-collection. But $(XY_1)(XY_2) = XY_1Y_2 = XY$, where $Y = Y_1Y_2$. Now Y is an element of H and so we have that $[XY]$ is an M-collection.

(1.5) *The intersection of any family of elements of an inductive M-collection is again an element of the collection.*

Let Z be the intersection of the collection $H = [Y]$ of elements of the inductive M-collection $G = [X]$. If G is empty the result is obvious. In view of (1.1) there is a minimal element, say Z_0, which contains Z. Now $Z \subset Z_0Y$ for each Y and since G is an M-collection we have $Z_0Y \in G$. Thus since Z_0 is a minimal

set we have $Z_0 \subset Z_0 Y$ and so Z_0 is contained in Y for each Y. Hence $Z_0 \subset Z$. Thus $Z = Z_0$ and since $Z_0 \in G$ this completes the proof.

(1.6) *If G is an inductive collection, then the intersection of any family of elements of $M(G)$ is an element of $M(G)$.*

This follows from (1.5) and (1.4.5).

If G is an inductive collection and A is any subset of S *which is contained in an element of* G, then we denote by $k(A)$ a (not necessarily unique) minimal element of G which contains the set A. Such an element will exist in view of (1.1). Since k is not single-valued it is to be understood that, when it occurs in the statement of a theorem, $k(A)$ refers to all possible minimal sets which contain A. From (1.4.2) and (1.6) there will exist a *unique* set $h(A)$ of $M(G)$ containing a given subset A of S.

(1.7) *If G is an inductive collection then*

(1.7.1) $$A \subset k(A) \subset h(A);$$

(1.7.2) *the closed set Z is in $M(G)$ if and only if $E \subset Z$ implies $k(E) \subset Z$;*

(1.7.3) $$h(h(A)) = h(A);$$

(1.7.4) $$h(A) + h(B) \subset h(A + B);$$

(1.7.5) $$h(A) \subset h(B) \text{ if } A \subset B;$$

(1.7.6) $$h(h(A) + h(B)) = h(A + B).$$

Proof. It is clear that $A \subset k(A)$ and $A \subset h(A)$. The second inclusion follows from (1.7.2) which we now prove. If $Z \in M(G)$ and Z contains E, then (assuming of course the existence of $k(E)$) we see that $Z \cdot k(E) \in G$ because $k(E) \in G$. Since $Z \cdot k(E) \subset k(E)$ and the latter is minimal we have $Z \cdot k(E) = k(E)$ and so $k(E) \subset Z$. Let us suppose that $E \subset Z$ implies $k(E) \subset Z$. Let $K \in G$. We have to show that $KZ \in G$. Since $KZ \subset K$ there must exist a set $k(KZ)$ which is necessarily contained in K. But for any such set $k(KZ) \subset Z$. Hence $k(KZ)$ is a subset of KZ. Hence we have $k(KZ) = KZ$ and so we infer that KZ is an element of G. The remaining assertions are readily proved.

The reader will notice the analogy between the operator h and the closure operator. This observation has also been made by J. W. T. Youngs for certain choices of G. The following examples may be of interest. Let S be the plane and G the collection of all closed line intervals including the unbounded ones, points and the null-interval. Here $M(G)$ is the class of all closed and convex sets. Again, let S be a Peano space and G the collection of all simple arcs, points and the null-set. Then $M(G)$ is the class of all A-sets (= closed arc-sets).

We use the symbol $X|Y$ to indicate that the sets X and Y are mutually separated, that is, $\overline{X}Y + X\overline{Y} = 0$. Let $F = [J]$ be a collection of closed subsets of S. We say that X and Y are F-separated if one of them is null or if there exists a J such that $S = P + Q$, $PQ = J$, $X \subset P$, $Y \subset Q$, $J(X + Y) = 0$, where P

and Q are closed. A set is F-connected if it is not the union of two non-null F-separated sets, see [20; §§4, 5]. If S is a normal space and F is the collection of all closed sets then a closed set is F-connected if and only if it is connected in the usual sense.

(1.8) *If each element of F is compact, then the property of being closed and F-connected is inductive.*

Proof. Let $[X]$ be an M-collection of closed and F-connected sets. Suppose that the intersection Y of this collection is not F-connected. Then there is a set J such that $S = P + Q$, $J = PQ$, $JY = 0$, $YP \neq 0 \neq YQ$, with P and Q closed. We assert that for some X we have $XJ = 0$. If not, then $[XJ]$ is an M-collection of non-null closed sets in the compact space J. By (1.3) we conclude that YJ is not void. Hence, for some X, $XJ = 0$, $XP \neq 0 \neq XQ$, a contradiction.

As indicated above, letting F be the collection of all closed sets in S, we have

(1.8.1) *In a compact normal space the property of being a continuum is inductive.*

It is clear that if $[B]$ is a collection of inductive properties and B_0 is the property of having all of the properties B, then B_0 is itself inductive.

(1.9) *If A and B are disjoint closed subsets of the compact normal space S and no continuum meets both A and B, then S is the union of disjoint closed sets containing A and B respectively.*

Proof. Let a and b be points of A and B respectively. Let A_0 , B_0 be the closures of a, b respectively. It will be recalled that our definition of normality does not require that a point be closed. Then A_0 and B_0 are disjoint continua. Let F be the family consisting precisely of the null-set. The property P of being F-connected and meeting both A_0 and B_0 is inductive by (1.8) and (1.3). Suppose that S has property P. Then by (1.2) there is a minimal closed set with property P. By an argument similar to that given by Moore [12; 18, Theorem 31] this set must be a continuum. This is contrary to our hypothesis. Hence $S = M + N$, $MN = 0$, $A_0M \neq 0 \neq B_0N$, M and N closed. Now, A_0 and B_0 being continua, we must have $A_0 \subset M$, $B_0 \subset N$. Since the sets M and N are also open, it follows readily from the compactness of S that the desired decomposition exists, the argument here being that of [12; 21, Theorem 35].

For compact metric spaces this and the following results are classical and we refer the reader to the books of Menger, Moore and Whyburn for historical and bibliographical data. For the sake of scientific accuracy we feel obliged to state these results, since sharp and unexpected differences arise in treating separable and non-separable spaces. Thus, in a compact metric locally connected space in which each point is either a cutpoint or an endpoint it is well known that every subcontinuum is unicoherent. Nevertheless a locally connected compact Hausdorff space may satisfy this condition and still contain a

continuum that is not unicoherent. And indeed, for a pre-assigned positive integer n, it is possible to construct such a space with a positive n-th Betti number. Moreover, it would be reasonable to expect that our results (1.11) and (2.10) would fail to hold in the absence of separability since the known proofs depend strongly on the second axiom of countability.

(1.10) *If S is a compact normal continuum and X is any subset of S then there exists a minimal continuum containing X. If the subsets X and Y of S are disjoint and closed then there exists a minimal continuum meeting both X and Y. If this continuum be denoted by K, then $K - X$, $K - Y$ and $K - (X + Y)$ are connected and $K(X + Y) \subset \overline{K - (X + Y)}$.*

A proof of this may be constructed using previous results and proofs analogous to those we have given. For the latter part see [12; 22]. From this result it follows that for any closed set Z of S and any component C of $S - Z$ we have $\overline{C}Z \neq 0$. Further, if X and Y are as in (1.10) then there is a component C of $S - (X + Y)$ whose closure meets both X and Y. For metric spaces E. W. Miller [11] has shown that there is a semi-continuum C with this property. So far we have not been able to get Miller's result in our present set-up. Sûra-Bûra has recently proved (under slightly less general conditions) certain results which are closely related to the latter part of (1.10) and which follow at once from this theorem.

Other examples of inductive properties may be given. Thus, let n be a positive integer and let B be the property that each n-cycle in X (a closed set) is homologous to 0 in X. Using a convergence theorem for cycles [7; VII, (15.1)] we infer that property B is inductive for compact normal spaces. This gives an extension to a result of Vietoris who proved analogous theorems for compact metric spaces.

(1.11) *If S is a compact connected T_1 space with at least two points, then S contains at least two non-cutpoints.*

We have to show that if p is any point of S, then there is a non-cutpoint q, distinct from p. We consider the class $[K]$ of all proper subcontinua of S which contain p. This collection is partially ordered by defining K_1RK_2 if $K_1 = K_2$ or $K_1 \subset K_2^0$, where X^0 is the interior of the set X. Since p is a continuum we know that there exists a maximal ordered subsystem P_0 of $[K]$. Let S_0 be the union of all the continua in P_0. Suppose that $S_0 = S$. We cannot have a largest element in P_0 since, for such a K, we would have $K = S$. It follows that each point s in S is contained in some set K^0 with $K \in P_0$. Since S is compact we can cover it by a finite family of elements of the collection P_0. Now P_0 is ordered and among this finite collection there is a largest set and we infer that for this K we have $S = K$, a contradiction. We conclude that there is a point q in $S - S_0$. Suppose that $S - q = U + V$, $U \mid V$, $p \in U$. Since S_0 is connected it must be a subset of U. Now $U + q$ is a continuum and $S_0 \subset (U + q)^0$. Thus each $K \subset (U + q)^0$, contrary to the fact that P_0 is maximal. This completes the proof.

Essentially the same proof may be used to extend a theorem, due to Eilenberg, of a similar nature. We have recently learned that R. L. Wilder has also found a proof of (1.11).

2. In this section we always assume that S is a *compact connected Hausdorff* space. If p and q are points of S then we write $p \sim q$ if there exists no point x such that $S - x = U + V$ with $p \in U$, $q \in V$ and $U \mid V$. In other words $p \sim q$ means that no point separates p and q in S. We have at once the following rules: $R_1 : p \sim p$; $R_2 : p \sim q$ implies $q \sim p$; R_3 : if $p \sim x \sim q$ and $p \sim y \sim q$ then $x \sim y$; R_4 : if $p \sim z \sim q$ then no point other than z can separate p and q in S.

By a chain we mean a subcontinuum X in S which satisfies the following chain condition: if p and q are distinct points of X and $p \sim x \sim q$, then $x \in X$. In a locally connected metric space a chain is an A-set [6] or a closed arc-set [2]. But our definition follows more closely that of Radó-Reichelderfer [13]. We give an example later to show that chains and J-sets [5] are distinct, as are chains and A-sets (as redefined by Whyburn [26]). Indeed, Kelley [5] has given an example to show that J-sets are not invariant under non-alternating transformations while chains are invariant.

If G is the class of all continua, then by a *semi-chain* is meant an element of $M(G)$. It is readily seen that chains and semi-chains are distinct classes. From earlier results we see that a semi-chain is a continuum and that the intersection of semi-chains is a semi-chain.

(2.1) *Each chain is a semi-chain.*

Proof. Let X be a chain and Y a continuum such that $XY = A + B$ where $A \mid B$. Let a and b be points of A and B respectively and let J be a minimal subcontinuum of Y containing $a + b$. It is clear that J is not contained in X. If $S - z = U + V$, $U \mid V$, $a \in U$, $c \in V(J - X)$, then $J - z = JU + JV$, $JU \mid JV$, $a \in JU$ and $c \in JV$. But then [12; 66, Theorem 91] we must have $b \in JV$. Now $X + JV$ is connected since XJV contains b and, as is readily verified, JV is a connected set. Also $X + JV$ contains $a + c$. Hence z cannot separate a and c in S, so that $a \sim c$. Similarly $b \sim c$. Hence by the chain condition we have c in X. This is a contradiction.

(2.2) *The intersection of any collection of chains is a chain.*

This follows at once from the definition, (2.1) and (1.4.3).

We write $C(A)$ for the intersection of all chains which contain the set A. If we denote by G the set of all chains then we have (as is readily verified) $M(G) = G$. The properties of the operator C may be at once deduced from (1.7). Here, of course, we have $h = k = C$. For each point s in S we write $E(s)$ for the set of all points x such that $x \sim s$. These sets are analogous to cyclic elements but are more closely related to a class of sets considered by Ayres [3] with which they are identical if S is locally connected and metric. See also [12] and [5].

(2.3) *Each set $E(s)$ is a chain.*

Proof. That $E(s)$ is a continuum follows at once by a known argument [12; 67, Theorem 93]. Let x and y be points of $E(s)$ and let $x \sim z \sim y$. We have $x \sim s \sim y$ and hence from R_3 we infer at once that $s \sim z$. But then $z \in E(s)$.

(2.4) *If a and b are distinct points and S_0 is the set of all points x such that $a \sim x \sim b$, then, if $a \sim b$, the set S_0 is a chain and no point separates any pair of points of S_0 in S.*

It is clear that $S_0 = E(a)E(b)$ and so by (2.3) and (2.2) the set S_0 is a chain. The second part may be proved as was the second part of (2.3).

(2.5) *For any point x in S such that $S - x = U + V$, $U \mid V$, the set $U + x$ is a chain.*

Clearly the set $U + x$ is a continuum. Let $a \in U$ and $b \in U + x$. Further, suppose that $a \sim y \sim b$ and that y is not in $U + x$. Since no point other than y can separate a and b in S and $a + b$ is contained in a continuum not containing y, it is clear that $a \sim b$. Let Y be the set $E(a)E(b)$. Now y is in V and, by (2.4), Y is a continuum meeting both U and V. Hence x is in Y. Thus x separates a and y in S, contrary to (2.4).

By a *prime chain* is meant a non-degenerate chain which contains no non-degenerate subset which is a chain.

(2.6.1) *Let $a \sim b$, $a \neq b$. Then the set $E(a)E(b)$ is a prime chain.*

(2.6.2) *If X is a prime chain and a and b are distinct points of X, then $a \sim b$ and $X = E(a)E(b)$.*

Proof. From (2.4) we see that X is a chain. Let Y be a non-degenerate chain contained in Y. If a_0 and b_0 are in Y and x is in X, then by (2.4) we have $a_0 \sim x \sim b_0$. Since Y is a chain, $x \in Y$ and so $X \subset Y$. For (2.6.2) let X be a prime chain. Suppose that a and b are points of X and that z separates a and b in S, say $S - z = U + V$, $U \mid V$, $a \in XU$, $b \in XV$. Now by (2.5) the set $U + z$ is a chain and hence so is $X(U + z)$. This latter is non-degenerate and is a proper subset of X. This contradicts the fact that X is a prime chain. We conclude that $a \sim b$. Now $E(a)E(b)$ is a chain and so X is contained in this set, since otherwise $XE(a)E(b)$ would be a non-degenerate proper subset of X. But $x \in E(a)E(b)$ means $a \sim x \sim b$ and hence $x \in X$. Thus $X = E(a)E(b)$.

In a semi-locally connected metric space [26] the sets X of (2.6) are the E_0 sets. In general, in a metric space they are the simple-links of Moore or equivalently the F-sets of Kelley. Now it is clear that if a prime chain meets a chain in two points then it is a subset of that chain. But this does not imply (even in a metric space) that each chain is a connected set of prime chains (i.e., a J-set of Kelley). Indeed, let Y be the unit interval, C the Cantor set on Y, and let S be Y together with the collection of unit intervals each perpendicular to Y at a point of C. Let p be a point of C which is a limit point both from the left and

from the right. If X is the interval at p, then X is a prime chain which meets Y in p, and p is neither an endpoint nor a cutpoint. Hence the chain Y contains part but not all of X. Now Whyburn [23; VIII, (6.3)] has shown that generalized A-sets are invariant under non-alternating transformations. The set Y is not such an A-set but is invariant under this type of transformation.

(2.7) *If X and Y are chains (semi-chains) and XY is not null then $X + Y$ is a chain (semi-chain).*

Proof. First we have the result for semi-chains. Let K be a continuum. Since we wish to show that $K(X + Y)$ is a continuum we may suppose that K meets both X and Y. Hence $K + X$ is a continuum and thus $(K + X)Y = KY + XY$ is also. Since these sets are non-null, KXY is also. Now $K(X + Y) = KX + KY$ and the summands are continua with at least one point in common. Hence $K(X + Y)$ is a continuum. Suppose now that X and Y are chains. Let p be a point such that $a \sim p \sim b$, where $p \in S - (X + Y)$, $a \in X$ and $b \in Y$. Clearly we have $a \sim b$. Let $Z = E(a)E(b)$. Then Z is a prime chain by (2.6). Since ZX is a chain we must have $ZX = a$. Similarly $ZY = b$. Now $X + Y$ is a semi-chain and we infer that $Z(X + Y) = a + b$ is a continuum. This contradiction completes the proof.

This result has an interesting dual:

(2.8.1) *If $X + Y$ and XY are semi-chains then so are X and Y, if they are closed.*

(2.8.2) *If X, Y and Z are semi-chains and XY, YZ and ZX are all non-null, then XYZ is non-null.*

Proof. Suppose that $XYZ = 0$. Then since $XY \neq 0$, $X + Y$ is a semi-chain and so $(X + Y)Z = XZ + YZ$ is a continuum. But XZ and YZ are disjoint non-null sets, a contradiction.

This last result and the preceding have not, so far as we are aware, been stated in the literature even in the classical case.

(2.9) *Let X be a semi-chain.*
(2.9.1) *If $S - X = U + V$, $U \mid V$, then $U + X$ is a semi-chain.*
(2.9.2) *If Q is a quasi-component of $S - X$, then $Q + X$ is a semi-chain.*
(2.9.3) *If Z is a component of $S - X$, then $Z + X$ is a semi-chain and $\overline{Z} - Z$ is a subcontinuum of X.*

Proof. Let K be a continuum. Then $K - KX = KU + KV$, $KU \mid KV$. Since K and KX are continua, $UK + XK = K(U + X)$ is also. For (2.9.2) we have that Q is the intersection of all possible sets Q', where $S - X = Q' + R'$, $Q' \mid R'$, Q contained in Q'. For each Q' the set $Q' + X$ is a chain. Hence the intersection of all these sets is a chain. But this intersection is $Q + X$. To prove (2.9.3) let J be a continuum minimal relative to the property of containing the points a and b. Suppose first that $a \in X$ and $b \in Z$. Then JX is a continuum containing a and so $J - X$ is connected and since it meets Z it must be con-

tained in Z; see [12; 79, Theorem 115]. Hence we infer that $J = (J - X) + JX$ is a subset of $Z + X$. Now it is easy to see that a semi-chain may be defined as a closed set which contains with each pair of points each continuum minimal relative to the property of containing these points; see (1.7.2). The proof is complete for this case. Suppose that a and b are both in Z. If $J - X$ is connected it is contained in Z and we argue as before. If it is not connected, then, since JX is connected, we see that $J - X$ is the union of two mutually separated connected sets each of which meets Z, and so it follows that J is contained in $Z + X$.

For additional references concerning the prototypes of chains and semi-chains see the papers of Ayres and Whyburn in the bibliography of [6]. For (2.10) see [5] and [6]. Both of the proofs depend on separability and so are inapplicable in our set-up.

(2.10) *If p is any point of S, then p is an endpoint, a cutpoint, or is contained in a unique prime chain.*

Proof. Suppose that p is neither an endpoint nor a cutpoint. We wish to exhibit a point which $\sim p$. We may clearly assume the existence of a point q such that $S = H + K$, $HK = z$, $H \neq z \neq K$, where $p \in H$, $q \in K$, so that H and K are continua. Let $[K']$ be a maximal ordered (by inclusion) collection of sets such that we have $S = H' + K'$, $H'K' = z'$, $H' \neq z' \neq K'$, with p in H' and H', K' continua. Let H_0 be the intersection of all the sets H' and K_0 the union of all the sets K'. It is clear that the collection $[H']$ is ordered and so by (1.8.1) H_0 is a continuum. Further, we see that K_0 is connected.

Observe first that there must be a point in H_0 distinct from p. For otherwise let $p = H_0$ and let U be an open set which contains p. If for each H' we have $(S - U)H'$ non-void, then by (1.3) it follows that $(S - U)H_0$ is also non-void. This is impossible and so we infer that some $H' \subset U$. Now $H' - z'$ is an open set with boundary z'. This being true for each neighborhood U, it follows that p is an endpoint.

Suppose that x and y are points of H_0K_0. Since $[K']$ is an ordered system we must have x and y in some set K'. But $x + y$ is contained in the corresponding H' and thus $x = y$. Let $x = H_0K_0$. Since p is in no set K' we cannot have $x = p$. Let us suppose that $x \nsim p$. In this case there is a point z such that $S - z = U + V$, $U \mid V$, $x \in U$ and $p \in V$. Since x, $p \in H_0$ and H_0 is connected, we have $z \in H_0$. And since H_0 and K_0 meet in x it follows that K_0 is contained in U. Thus each set K' is a proper subset of the continuum $U + z$ and we infer that the system $[K']$ is not maximal, a contradiction. Hence $x \sim p$ or $H_0K_0 = 0$. We make the latter assumption. From the connectedness of S and the fact that H_0 is closed there must be a point in the set $\overline{K}_0H_0$. Let p be a point of this set and suppose first that it contains another point, say w. The set $K_0 + p + w$ is connected, as is H_0. Thus if $p \nsim w$ the point which separates them must lie in both of these sets. But this is impossible since $H_0K_0 = 0$. We conclude that $p = \overline{K}_0H_0$. But from this it is seen that (as in the argument above),

because $p \neq H_0$, p must be a cutpoint. It follows that p is not in $\overline{K}_0 H_0$. Let f be a point of this set and suppose that $f \sim p$. Then there is a point z such that $S - z = U + V$, $U \mid V$, $f \in U$, $p \in V$. Since $f \in \overline{K}_0$ it is clear that U meets K_0. Further $f + p \subset U$ and so z is contained in $S - K_0$. Since K_0 is connected it is certainly in U. As in the argument above we infer that $[K']$ is not maximal. Accordingly we have $f \sim p$.

Now let X be the prime chain $E(f)E(p)$. Suppose that p is contained in another prime chain X_0. Let g be a point of X_0 distinct from p and f. By (2.6.2) we have $g \sim p$. Hence $g \sim p \sim f$. Since p is a non-cutpoint it follows by R_4 that $p \sim g \sim f$. Hence $g \in X$ and so $X_0 \subset X$. Thus $X_0 = X$. This completes the proof of (2.10).

It is worthwhile to point out the relation between the situation as we consider it and that in metric spaces. In the latter instance a simple-link may be defined as a set M_p consisting of a non-cutpoint p together with all points x with $x \sim p$. Now it is not hard to see that

(2.11) *Each set M_p is a prime chain.*

For $M_p = E(p)$ and so is a chain. Also p is contained in a unique prime chain P and, if $q \in P$, then $q \sim p$ and so by (2.6.1) $E(p)E(q) = P$. Since P is a prime chain meeting the chain M_p in two points we have $P \subset M_P$. But if $r \in M$, $p \neq r \neq q$, then $r \sim p$ and since $E(r)E(p)$ is a prime chain we have r in P by (2.10). Hence $P = M_p$.

The converse of (2.11) is false, and indeed there exists a locally connected space S in which every point is either an endpoint or a cutpoint and which certainly contains no non-degenerate set M_p but which does contain a prime chain. In metric spaces the converse of (2.11) is true. From (2.2) and (2.10) it follows without difficulty that

(2.12) *If the intersection of distinct prime chains is not null, then it is a cutpoint.*

3. For the present M and N will denote T_1-spaces and Δ will be a single-valued transformation of M onto N, $\Delta M = N$. The correspondence Δ will not necessarily be continuous. For the next result for metric spaces see [16].

(3.1) *If M and N are compact Hausdorff spaces, Δ is continuous if and only if*

(T) *the sets Y_1 and Y_2 of N are separated if and only if $\Delta^{-1}Y_1$ and $\Delta^{-1}Y_2$ are separated.*

Proof. Let Δ be continuous. Then for each set Y of N we know that $\overline{\Delta^{-1}Y}$ is contained in $\Delta^{-1}\overline{Y}$ and from this it is readily seen that $Y_1 \mid Y_2$ implies $\Delta^{-1}Y_1 \mid \Delta^{-1}Y_2$. Also (the spaces concerned being compact) it is known that $\overline{\Delta X} = \Delta\overline{X}$, and from this we draw the reverse implication. Suppose now that condition (T) holds. We need only show that $\Delta\overline{X} \subset \overline{\Delta X}$. Let y be a point of the former set. Then, since $\Delta^{-1}y$ meets $\overline{X}$, we infer that X and $\Delta^{-1}y$ are non-separated. It

follows that the latter set is not separated from $\Delta^{-1}\Delta X$. From this it may be concluded that y and ΔX are non-separated. Because N is a Hausdorff space, y is a closed set and so $y \in \overline{\Delta X}$.

(3.2) *Let $G = [X]$ be an M-collection of closed subsets of M with intersection X_0. Then, if $\Delta^{-1}y$ is compact for each $y \in N$, we have $Y_0 = \Delta X_0$, where Y_0 is the intersection of the sets $[\Delta X]$.*

Proof. It is readily seen that ΔX_0 is contained in Y_0, so that we have to prove the reverse inclusion. Let $y \in Y_0$. Since G is an M-collection so also is the collection $[X\Delta^{-1}y]$ and each set is non-vacuous and closed (in $\Delta^{-1}y$). This latter set being compact it is manifest that there is a point common to all the sets $[X\Delta^{-1}y]$. Denoting this point by x we then have $y = \Delta x$ contained in ΔX_0.

In connection with the next result, see [7; II, (5.5)], [12], and for metric spaces [21].

(3.3) *Let N be a compact space and let Δ be closed. Then, if for each $y \in N$ the set $\Delta^{-1}y$ is compact, the space M is also compact.*

Proof. In view of (1.3) we have to show that if $G = [X]$ is an M-collection with intersection X_0 then X_0 is not void. Let H be the collection whose elements are finite intersections of sets $[\Delta X]$. Then H is an M-collection with the same intersection as $[\Delta X]$, say Y_0. Since N is compact Y_0 is not void. By (3.2) $\Delta X_0 = Y_0$ and we infer that X_0 is not void.

From this point on we assume that Δ satisfies the condition (T) *of* (3.1). In general we do not assume that M and N are compact.

DEFINITION. If $G = [Z]$ is a collection of closed sets covering N, then Δ is said to be monotone relative to the collection G (or G-monotone) provided that if $Z \in G$ and $M - \Delta^{-1}Z = M_1 + M_2$, $M_1 \mid M_2$, then no set $\Delta^{-1}y$ meets both M_1 and M_2.

We say (with Whyburn [24]) that Δ is non-alternating if it is monotone relative to the set of all points of N. Various modifications of non-alternating transformations have been suggested by Hall and Schweigert, Odle, Vance and Wardwell. Further, Δ is said to be monotone if $\Delta^{-1}y$ is connected for each $y \in N$.

Let M be the circle $|z| = 1$ in the complex plane and let $\Delta z = x$, $z = x + iy$, so that Δ is merely the projection of M onto the segment of the real axis from -1 to $+1$, which is the space N. Then Δ is monotone relative to any collection of continua which cover N, but Δ is not monotone relative to any collection which contains a disconnected set.

Let M be as before and let $\Delta z = z^2$. Here Δ is not monotone relative to any collection which contains a proper closed subset of N. As we show later a monotone transformation is monotone relative to any collection.

(3.4) *In order that* $\Delta M = N$ *be G-monotone it is necessary and sufficient that for each* $Z \in G$ *and each decomposition*

$$M - \Delta^{-1}Z = M_1 + M_2 , \qquad M_1 \mid M_2 \tag{3.4.0}$$

we have

$$M_i = \Delta^{-1}M_i \qquad (i = 1, 2); \tag{3.4.1}$$

$$\Delta M_1 \mid \Delta M_2 ; \tag{3.4.2}$$

$$\Delta M_1 \cdot \Delta M_2 = 0; \tag{3.4.3}$$

$$\Delta^{-1}\Delta M_1 \cdot \Delta^{-1}\Delta M_2 = 0. \tag{3.4.4}$$

Proof. Let Δ be G-monotone. If $x \in M_1$ we must have that $\Delta^{-1}\Delta x \subset M_1$ and this clearly implies $M_1 = \Delta^{-1}\Delta M_1$. Now (3.4.2) follows from (3.4.1) and condition (T); (3.4.3) is an obvious consequence of (3.4.2). Clearly (3.4.3) implies (3.4.4). Suppose that (3.4.4) holds. Then if the inverse of a point meets both M_1 and M_2 it must lie in the set $\Delta^{-1}\Delta M_1 \cdot \Delta^{-1}\Delta M_2$. We infer that Δ is G-monotone.

We proceed now to justify the term "relatively monotone".

(3.5) *If* M *and* N *are compact Hausdorff spaces then in order that* Δ *be monotone it is necessary and sufficient that it be monotone relative to the collection of all closed sets in* N.

Proof. It is readily seen that if Δ is monotone it is monotone relative to the collection of all closed sets in N. Suppose now that there is a point n of N such that $\Delta^{-1}n = N_1 + N_2 , N_1 \mid N_2$. Clearly N_1 and N_2 are closed and so since M is normal there exists a closed set F such that $M - F = M_1 + M_2 , M_1 \mid M_2 ,$ $N_1 \subset M_1$ and $N_2 \subset M_2$. Since M is compact we know that Δ is closed so that $\Delta^{-1}\Delta F$ is closed. Hence we have $M - \Delta^{-1}\Delta F = (M_1 - \Delta^{-1}\Delta F) + (M_2 - \Delta^{-1}\Delta F)$ $= M_1' + M_2' , M_1' \mid M_2'$. Clearly $\Delta^{-1}n$ meets both M_1' and M_2' .

(3.5.1) *Let* M *and* N *be compact locally connected Hausdorff spaces and let* G *be the collection of all closed subsets of* N *each of which has only a finite number of components. Then* Δ *is monotone if and only if it is G-monotone.*

The proof does not differ essentially from that for (3.5). We have merely to replace F by F_0 , a set with only a finite number of components. This is easily done since F is compact and M is locally connected.

To avoid circumlocution it is necessary to establish certain conventions. If S_0 is a subset of S_1 the X will be said to *cut* S_0 *in* S_1 if $S_1 - X = U + V$, $U \mid V$ and $S_0U \neq 0 \neq S_0V$. If, in addition, $S_0X = 0$ we say that X *separates* S_0 *in* S_1 . We omit "in S_1" if S_1 is the space under consideration.

(3.6.1) *If* Δ *is G-monotone, then* Z *cuts* N *if and only if* $\Delta^{-1}Z$ *cuts* M.

This follows at once from (3.4).

(3.6.2) *If M is connected, Δ is G-monotone and some component of $\Delta^{-1}Z$ cuts M then Z cuts N. In particular, if each set $\Delta^{-1}Z$ is totally disconnected, then the image of a cutpoint is a cutpoint.*

Proof. Let Y be the component of $\Delta^{-1}Z$ which cuts M so that $M - Y = M_1' + M_2'$, $M_1' \mid M_2'$. Since M and Y are connected, so is the set $M_i' + Y$, $i = 1, 2$. Neither set can be contained in $\Delta^{-1}Z$ since Y is a component of this set. Hence, if we put $M_i = M_i' - \Delta^{-1}Z$ we obtain a decomposition (3.4.0) and since $N - Z = \Delta M_1 + \Delta M_2$ it follows by (3.4.2) that Z cuts N. If p is a cutpoint, then p is contained in some set $\Delta^{-1}Z$ and is a component of this set if it is totally disconnected.

We may show as in (3.6.2) that

(3.6.3) *If M is connected, Δ is G-monotone, each set $\Delta^{-1}Z$ is totally disconnected and the set $\Delta^{-1}Z_0$ contains a closed set which irreducibly cuts M, then Z_0 cuts N.*

We now extend certain results due to G. T. Whyburn [23; VIII].

(3.7) *The following statements are equivalent:*
(3.7.0) *The transformation Δ is G-monotone.*
(3.7.1) *If the set $\Delta^{-1}Z$ separates p and q in M, then Z separates Δp and Δq in N.*
(3.7.2) *If Q is a quasi-component of $N - Z$, then $\Delta^{-1}Q$ is a quasi-component of $M - \Delta^{-1}Z$.*

Proof. (3.7.0) *implies* (3.7.1). By assumption, we have a decomposition (3.4.0) with $p \in M_1$ and $q \in M_2$. Hence $N - Z = \Delta M_1 + \Delta M_2$ with $\Delta p \in \Delta M_1$ and $\Delta q \in \Delta M_2$. The result follows by (3.4.2). (3.7.1) *implies* (3.7.2). Let Q be a quasi-component of $N - Z$ and suppose that there exists a decomposition (3.4.0) such that $\Delta^{-1}Q$ meets M_1 and M_2 in the points p_1 and p_2 respectively. Then since $\Delta^{-1}Z$ separates p_1 and p_2 we know that Z separates Δp_1 and Δp_2 contrary to the fact that Δp_1 and Δp_2 lie in the same quasi-component of $N - Z$. We may thus suppose that $\Delta^{-1}Q$ is contained in a quasi-component Q' of $M - \Delta^{-1}Z$. From the condition (T) we infer that $\Delta Q'$ is contained in a quasi-component of $N - Z$ and so $\Delta Q'$ is contained in Q. It follows that $Q' = \Delta^{-1}Q$. (3.7.2) *implies* (3.7.0). Let $n \in N$, $Z \in G$ and suppose that Q is the quasi-component of $N - Z$ which contains n. Then $\Delta^{-1}n$ is contained in $\Delta^{-1}Q$ and since this latter set is a quasi-component of $M - \Delta^{-1}Z$ it follows that $\Delta^{-1}Z$ does not separate $\Delta^{-1}n$ in M.

It is clear that when Δ is G-monotone and Q is a quasi-component of $M - \Delta^{-1}Z$ then $Q = \Delta^{-1}\Delta Q$. *From now on we assume that M and N are compact connected Hausdorff spaces.* The condition (T) will then be satisfied if we assume, as we shall, that Δ is continuous. One can easily give an example to show that, although no component of Z cuts N, nevertheless $\Delta^{-1}Z$ may cut M (cf. the result (3.6.2)). However, we do have

(3.8) *Let M and N be locally connected and suppose that each set $Z \in G$ has only a finite number of components no one of which cuts N. Then if Δ is G-monotone no set $\Delta^{-1}Z$ cuts M into more than a finite number of components.*

Proof. Let $A_1 + \cdots + A_n$ be a componentwise decomposition of Z. Since N is normal we can find pairwise disjoint open sets U_i such that $A_i \subset U_i$. Let $[R]$ be the collection of all components of $N - Z$. The sets $[R]$ together with the sets $[U_i]$ form an open covering of N. Because N is compact it is covered by a finite collection of these sets. By suitably adjusting the notation we may suppose that $N = R_1 + \cdots + R_m + U_1 + \cdots + U_s$. Suppose that the sets R_i do not cover $N - Z$ and let R_0 be a component of this set containing a point in no set R_i. Since N is connected, clearly the closure of R_0 meets Z and hence some component of Z, say A_j. Now $N - A_j = R_0 + (N - A_j - R_0)$ is connected and so, since R_0 is open, $\overline{R}_0$ meets $N - A_j - R_0$. But since $\overline{R}_0 - R_0$ is contained in Z we see at once that $\overline{R}_0$ meets some A_k, $k \neq j$. Hence R_0 meets both of the sets U_j and U_k. Since R_0 is connected it cannot be in $\sum U_i$ and so R_0 meets some set R_i, a contradiction. The sets R_i are thus the components of $N - Z$ and now the proof may be completed using (3.7).

We now give further conditions under which Δ will be monotone. These conditions involve both M and N and the collection G.

(3.9) *Let Δ be monotone relative to the collection of all finite subsets of N. Then*
(3.9.1) *if M is a regular curve Δ is monotone and N is a regular curve,*
(3.9.2) *if N is a regular curve Δ is monotone.*

Proof. First we shall prove (3.9.1). We need only show that Δ is monotone. Suppose that for some point n in N we have $\Delta^{-1}n = P + Q$, where P and Q are disjoint and closed. Since M is a regular curve we can find a finite set X such that $M - X = P' + Q'$, where $P' \mid Q'$ and $P \subset P'$, $Q \subset Q'$. Now $Y = \Delta X$ is finite and since $\Delta^{-1}n$ does not meet $\Delta^{-1}Y$ we have $M - \Delta^{-1}Y = P'' + Q''$, where $P'' = P' - \Delta^{-1}Y$, $Q'' = Q' - \Delta^{-1}Y$ and $\Delta^{-1}n$ meets both of these sets. To prove (3.9.2) we assume as before that $\Delta^{-1} = P + Q$, $P \mid Q$, so that, since M is normal, we can find a closed set X such that $M - X = P' + Q'$ with $P \subset P'$ and $Q \subset Q'$. Since n is not contained in ΔX we can find an open set U which contains n and whose boundary is a finite set which does not meet ΔX. We thus have $N - F(U) = R + S$, where $R \mid S$, $n \in R$ and $\Delta X \subset S$. Hence $M - \Delta^{-1}F(U) = \Delta^{-1}R + \Delta^{-1}S$ with $\Delta^{-1}n \subset \Delta^{-1}R$ and $\Delta^{-1}\Delta X \subset \Delta^{-1}S$. Since X does not meet $\Delta^{-1}R$ it follows that $M - \Delta^{-1}F(U) = P'\Delta^{-1}R + (Q' \cdot \Delta^{-1}R + \Delta^{-1}S)$. Since these summands are separated and $\Delta^{-1}n$ meets both of them we have a contradiction.

The result (3.9.2) was also discovered by P. A. White. For other results of his in connection with G-monotone transformations see [22]. There are certain generalizations of (3.9) which concern rational curves, etc. We state (without proof) only one of these.

(3.9.3) *If Δ is monotone relative to the collection of all subcontinua of N where M and N are locally connected and M is unicoherent, then Δ is monotone and N is unicoherent.*

(3.10) *If Δ is non-alternating and $M = A + B$ with $AB = x \in M$ with A and B closed, then $N = \Delta A + \Delta B$ with $\Delta A \cdot \Delta B = \Delta x$.*

Proof. Let $y = \Delta x$ and suppose that $y \neq z \in \Delta A \cdot \Delta B$. Now $M - \Delta^{-1}y = (A - \Delta^{-1}y) + (B - \Delta^{-1}y)$. Clearly these sets are separated. Since $z \in \Delta A$ it follows that $\Delta^{-1}z$ meets the first one and since $z \in \Delta B$ we see that $\Delta^{-1}z$ meets the second one. But then Δ fails to be non-alternating.

Modifying (in a very minor fashion) a definition due to G. T. Whyburn we say that a closed set E of S is a *nodal set* if $F(E) = E\overline{S - E}$ is a single point. If S is connected it is clear that E and $\overline{S - E}$ are connected.

(3.11) *If A is a nodal set in M and Δ is non-alternating, then ΔA is a nodal set in N.*

Proof. We have $M = A + \overline{M - A}$ and if we set $M - A = B$ we have $A\overline{B} = x$, a point. Hence $N = \Delta A + \overline{\Delta B}$ (since Δ is closed) and $\Delta A\overline{\Delta B} = \Delta x$, by (3.10). We want to show that $\overline{B} = \overline{N - \Delta A}$. Now $N - \Delta A \subset \Delta B$ and so we have $\overline{N - \Delta A} \subset \overline{\Delta B}$. Let y be a point of $\overline{\Delta B}$ distinct from Δx and let us suppose that y is not in $\overline{N - \Delta A}$. Then there is an open set U which contains y and such that $U\overline{N - \Delta A} = 0$; hence $U \subset \Delta A$. There exists an open set V with $y \in V \subset U$ and such that $V\Delta x = 0$. But since $\Delta x = \Delta A\overline{\Delta B}$ this implies that $V\overline{\Delta B} = 0$ so that y is not in $\overline{\Delta B}$, a contradiction. Hence we infer that $\Delta\overline{B} - \Delta x \subset \overline{N - \Delta A}$. Hence $\Delta\overline{B} \subset \overline{N - \Delta A} + \Delta x \subset \overline{\Delta B}$. Since $\overline{B}$ is a continuum and $\overline{\Delta B} = \Delta\overline{B}$ we conclude that $\Delta x \in \overline{N - \Delta A}$ and this completes the proof.

(3.12) *Let Δ be non-alternating and suppose that $a \sim b$ in N. Then if K is a continuum which meets both $\Delta^{-1}a$ and $\Delta^{-1}b$ there exist points a_0 in $K\Delta^{-1}a$ and b_0 in $K\Delta^{-1}b$ such that $a_0 \sim b_0$.*

Proof. Let J be a continuum in K minimal relative to the property of meeting both $\Delta^{-1}a$ and $\Delta^{-1}b$. The set $J_0 - \Delta^{-1}(a + b)$ is connected and so is $a_0 + J_0 + b_0$, where $a_0 \in J\Delta^{-1}a$ and $b_0 \in J\Delta^{-1}b$. Suppose that $M = A + B$ with $AB = x$, $a_0 \in A$ and $b_0 \in B$, $a_0 \neq x \neq b_0$. Since $x \in J_0$ we conclude that $a \neq \Delta x \neq b$. Hence by (3.11) we cannot have $a \sim b$. This completes the proof.

We now give our extension of Schweigert's theorem [14]. We have already mentioned Whyburn's generalization of this result [23; VIII, (6.3)].

(3.13) *If Δ is non-alternating and X is a chain in M, then ΔX is a chain in N.*

Proof. Let $Y = \Delta X$. To show that Y is a chain we must prove that if P is a prime chain which meets Y in the points a and b then $P \subset Y$. Now X is a continuum which meets both $\Delta^{-1}a$ and $\Delta^{-1}b$. Let a_0 and b_0 be the points given by (3.12). Further, let I be a continuum minimal relative to the properties (i) $a_0 + b_0 \subset I$ and (ii) $P \subset \Delta I$. Suppose that $M = A + B$, $AB = x$, $I(A - x) \neq 0 \neq I(B - x)$. Then since $a_0 \sim b_0$ we may assume that $a_0 + b_0 \subset A$. Now $N = \Delta A + \Delta B$ and $\Delta A \cdot \Delta B = \Delta x$. Further, $P \subset \Delta A$. Now by (2.5) IA is a continuum. Also $P \subset \Delta I = \Delta IA + \Delta IB \subset \Delta IA + \Delta I \cdot \Delta B \subset \Delta I$. From this we get, since $\Delta I \cdot \Delta B \cdot \Delta IA \subset \Delta A \cdot \Delta B = \Delta x$, $\Delta I = \Delta IA + \Delta x$. But ΔI is a continuum and so $\Delta I = \Delta IA$. But IA is a proper subcontinuum of I and IA has properties (i) and (ii). From this contradiction we conclude that no point

separates any pair of points of I in M. Hence if $c_0 \in I$ we have $a_0 \sim c_0 \sim b_0$ and so $c_0 \in X$. Thus $I \subset X$ and hence, since $P \subset \Delta I$, we have $P \subset Y$.

The following form of this theorem was stated in an abstract of this paper [20].

(3.14) *If Δ is non-alternating and A is a subset of M, then $C(\Delta A) \subset \Delta C(A)$.*

We wish to show (in a somewhat general fashion) the equivalence of theorems like (3.13) and (3.14). To this end suppose that there is defined for each set (whether in M or in N) a set function D such that $A \subset D(A)$, $D(D(A)) = D(A)$, and further such that $A \subset B$ implies $D(A) \subset D(B)$. We say that a set has *property D* if $D(A) = A$. By an *invariance theorem* we mean a theorem which asserts that (if Δ is a single-valued transformation from M onto N) whenever X has property D so also has ΔX. By a *covering theorem* we mean a theorem which asserts that for any set X we have $D(\Delta X) \subset \Delta D(X)$. What we wish to show is that *covering and invariance theorems are equivalent.* Suppose that the invariance condition holds. Since $A \subset D(A)$ we get $\Delta A \subset \Delta D(A)$ and $\Delta D(A)$ has property D (since $D(D(A)) = D(A)$ any set $D(A)$ has property D). Hence $D(\Delta A) \subset D(\Delta D(A)) = \Delta D(A)$. We want to show now that if the covering condition is valid so also is the invariance condition. Suppose that $\Delta X = Y$ where X has property D. Now $D(\Delta X) = D(Y)$ and since $D(X) = X$ we have $\Delta D(X) = \Delta X = Y$. Thus since $D(\Delta X) \subset \Delta D(X)$ we get $D(Y) \subset Y \subset D(Y)$ which proves invariance. This shows that (3.13) and (3.14) are equivalent.

It is known [23; VIII] that, when Δ is non-alternating and M and N are metric, then, if B is a simple-link in N, there exists a unique simple-link A in M such that $B \subset \Delta A$. This is no longer true when M is not metric. Indeed, there exists a locally connected space M which contains no non-degenerate simple-link and a monotone transformation of M onto a space N which is itself a simple-link. In this example N is a simple closed curve and is metric. The analogue of this theorem for prime chains is, however, true.

(3.15) *Let Δ be G-monotone, and let K be a closed subset of N which is not cut in N by any set KZ, $Z \in G$. If H is a closed set minimal relative to the property of mapping onto K, $\Delta H = K$, then no set $H\Delta^{-1}Z$ cuts H in M.*

Proof. Suppose the theorem is not true and let

(a) $M - W = M_1' + M_2'$, $M_1' \mid M_2'$, $HM_1' \neq 0 \neq HM_2'$, $W = H\Delta^{-1}Z$.

Let $Z_0 = \Delta W = KZ$. It is clear that we have

(b) $M - \Delta^{-1}Z_0 = M_1 + M_2$, $M_i = M_i' - \Delta^{-1}Z_0$, $M_1 \mid M_2$.

Hence

(c) $N - Z_0 = N_1 + N_2$, $N_i = \Delta M_i$, $N_1 \mid N_2$,

from (3.4). Thus $K - Z_0 = KN_1 + KN_2$ and since K is not cut in N by any element of G we may suppose that $K - Z_0$ is contained in N_1. From this

(d) $$\Delta^{-1}(K - Z_0) = \Delta^{-1}K - \Delta^{-1}Z_0 \subset \Delta^{-1}N_1 = M_1 ,$$

again from (3.4). Now we also have

(e) $$H\Delta^{-1}(K - Z_0) + W \subset H(M_1 + W) \subset H(M_1' + W) = H_0 ,$$

and this latter is a proper closed subset of H. But also,

(f) $$K = \Delta H = (K - Z_0)\Delta H + \Delta W \subset \Delta H_0 .$$

Hence we have H_0 mapping onto K, a contradiction.

(3.16) *Let Δ be non-alternating and K a subcontinuum in N which is not cut in N by any point. If H is a subcontinuum minimal relative to the property of mapping over K, $K \subset \Delta H$, then no point of H cuts H in M.*

We follow the proof of (3.15). In the present situation W is a point and so also is each element of G. Thus $Z_0 = Z$. Now the proof is valid up to the statement following (e), provided certain quite obvious changes are made. For the phrase following (e) we must have "and this latter is a proper subcontinuum of H". But from (2.5) it follows that $M_1' + W$ is a semi-chain and H being a continuum so is $H(M_1' + W)$. The remainder of the proof follows as in (3.15).

In view of the results of §1 it will be seen that sets H of the two preceding theorems will exist. It is also clear that no use was made of our standing hypothesis that M and N be compact Hausdorff spaces so that the results are valid for any T_1-spaces. The idea underlying the proof is due to G. T. Whyburn.

Suppose now that the situation is as in (3.16) and that K is a prime chain. The existence of a minimal set such as H being assured, we see, by reference to (2.6), that H is contained in a prime chain. It may be shown that H is unique (see [23]).

(3.17) *If Δ is non-alternating and B is a prime chain in N, then there is a unique prime chain A in M whose image covers B.*

(3.18) *If Δ is non-alternating it is monotone relative to the collection of all sets $[E(p)]$, $p \in N$.*

Proof. If not, there is a set $C = E(p) \subset N$ and a point n in N such that

$$M - \Delta^{-1}C = M_1 + M_2 , M_1 \mid M_2 , \qquad M_1\Delta^{-1}n \neq 0 \neq M_2\Delta^{-1}n.$$

Since $n \in N - C$ there exists a point z which separates n and p in N;

$$N - z = N_1 + N_2 , \qquad N_1 \mid N_2 , \qquad C - z \subset N_2 , \qquad n \in N_1 ,$$

from the definition of C as the set of all points y such that $y \sim p$. It follows that $\Delta^{-1}n \subset \Delta^{-1}N_1$, $\Delta^{-1}C - \Delta^{-1}z \subset \Delta^{-1}N_2$ and

$$M - \Delta^{-1}z = \Delta^{-1}N_1 + \Delta^{-1}N_2, \qquad \Delta^{-1}N_1 \mid \Delta^{-1}N_2,$$

and $\Delta^{-1}n$ meets each of the sets $M_1\Delta^{-1}N_1$ and $M_2\Delta^{-1}N_1$. But then

$$M - \Delta^{-1}z = M_1\Delta^{-1}N_1 + (M_2\Delta^{-1}N_1 + \Delta^{-1}N_2)$$

and the two summands of the right member of this equality are separated. Thus $\Delta^{-1}z$ separates $\Delta^{-1}n$ in M, a contradiction.

(3.19) *If $\Delta M = N$ is non-alternating, where M and N are locally connected and metric, then Δ is monotone relative to the collection of all chains of N.*

Proof. This follows the lines of that just given where z is replaced by the boundary of the complementary domain (which contains z) of the chain C. This boundary is a single point.

For the definition of a tree, see [17]. If a tree is a metric space then it is a dendrite (or acyclic continuous curve) in the usual sense of the term, see [18].

(3.20) *If $\Delta M = N$ is non-alternating and N is a tree, then Δ is monotone relative to the collection of all continua of N.*

Proof. What we have to show is that every subcontinuum of a tree is a chain. It will follow from the proof of this that if C is a chain then the complementary domain boundaries are single points. Hence the modification of the proof of (3.18) suggested for (3.19) will be available. Let a and b be any points of N. Since N is locally connected and normal there is a closed set Z which is minimal relative to the property of separating a and b in N. Further, Z is the common boundary of A and B, the components of $N - Z$ containing a and b. Suppose that Z contains two points p and q. For each point t of $N - (a + b + Z)$ we can find an open connected set $U(t)$ which lies in this set. For any point t of one of the sets a,b,Z we can find an open connected set $U(t)$ which does not meet the remaining two sets. We may also assume that $U(p)$ and $U(q)$ do not meet. Let $[V]$ be the finite open covering in (i) of [17] which refines the covering $[U(t)]$. It is easy to see (as in the argument given for (A_1) of [17]) that $[V]$ contains a simple closed chain. This is impossible from the definition of $[V]$. Thus Z is a point and so any pair of points of N can be separated by a single point. Hence [12; 74, Theorem 105] it follows that if X is a continuum and x is a point not in X then x can be separated from X by a point. Now if a and b are points of X then we cannot have $a \sim x \sim b$. Hence X is a chain, the chain condition being vacuously satisfied.

We say that Δ is quasi-monotone provided that, if R is an open connected subset of N and R_1 is a component of $\Delta^{-1}R$, then $R = \Delta R_1$. Modifying in a minor way a recent result of Whyburn [25], we see that if N is a tree, Δ is quasi-

monotone and further $\Delta^{-1}e$ is connected for each endpoint e of N, then Δ is non-alternating. In view of (3.20) we then have

(3.21) *If Δ is quasi-monotone and $\Delta^{-1}e$ is connected for each endpoint e of N, where M is locally connected and metric and N is a tree, then Δ is monotone relative to the collection of all continua of N.*

Of course, here, Δ carries chains into chains. It is also to be observed that there are other types of mappings which preserve chains [19]. However, it seems difficult to characterize more directly the class of all such transformations.

Bibliography

1. Alexandroff-Hopf, *Topologie*, Berlin, 1935.
2. W. L. Ayres, *Concerning the arc-curves and basic sets of a continuous curve, second paper*, Transactions of the American Mathematical Society, vol. 31(1929), pp. 595–612.
3. W. L. Ayres, *On the structure of a plane continuous curve*, Proceedings of the National Academy of Sciences, vol. 13(1927), pp. 749–754.
4. F. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig, 1914.
5. J. L. Kelley, *A decomposition of compact continua and related theorems on fixed sets under continuous transformations*, Proceedings of the National Academy of Sciences, vol. 26(1940), pp. 192–194.
6. C. Kuratowski and G. T. Whyburn, *Sur les éléments cycliques et leurs applications*, Fundamenta Mathematicae, vol. 16(1930), pp. 305–331.
7. S. Lefschetz, *Algebraic Topology*, American Mathematical Society Colloquium Publications, vol. XXVII, New York, 1942.
8. R. G. Lubben, *Concerning the decomposition and amalgamation of points, upper semi-continuous collections, and topological extensions*, Transactions of the American Mathematical Society, vol. 49(1941), pp. 410–466.
9. K. Menger, *Kurventheorie*, Leipzig, 1932.
10. A. N. Milgram, *Partially ordered sets and topology*, Proceedings of the National Academy of Sciences, vol. 26(1940), pp. 291–293.
11. E. W. Miller, *Some theorems on continua*, Bulletin of the American Mathematical Society, vol. 46(1940), pp. 150–157.
12. R. L. Moore, *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, vol. XIII, New York, 1932.
13. T. Radó and P. Reichelderfer, *Cyclic transitivity*, this Journal, vol. 6(1940), pp. 474–485.
14. G. E. Schweigert, *Concerning non-alternating interior transformations*, (abstract), Bulletin of the American Mathematical Society, vol. 44(1938), p. 636.
15. J. W. Tukey, *Convergence and Uniformity in Topology*, Princeton, 1940.
16. A. D. Wallace, *Concerning relatively non-alternating transformations*, Proceedings of the National Academy of Sciences, vol. 27(1941), pp. 182–185.
17. A. D. Wallace, *A fixed point theorem for trees*, Bulletin of the American Mathematical Society, vol. 47(1941), pp. 757–760.
18. A. D. Wallace, *Monotone coverings and monotone transformations*, this Journal, vol. 6(1940), pp. 31–37.
19. A. D. Wallace, *0-regular transformations*, American Journal of Mathematics, vol. 62(1940), pp. 277–284.
20. A. D. Wallace, *Separation spaces*, Annals of Mathematics, vol. 42(1941), pp. 687–697.
21. A. D. Wallace, *Some characterizations of interior transformations*, American Journal of Mathematics, vol. 61(1939), pp. 757–763.

22. P. A. White, *On certain relatively non-alternating transformations*, (abstract), Bulletin of the American Mathematical Society, vol. 46(1940), p. 435.
23. G. T. Whyburn, *Analytic Topology*, American Mathematical Society Colloquium Publications, vol. XXVIII, New York, 1942.
24. G. T. Whyburn, *Non-alternating transformations*, American Journal of Mathematics, vol. 56(1934), pp. 294–302.
25. G. T. Whyburn, *A relation between non-alternating and interior transformations*, Bulletin of the American Mathematical Society, vol. 46(1940), pp. 320–321.
26. G. T. Whyburn, *Semi-locally connected sets*, American Journal of Mathematics, vol. 61(1939), pp. 733–749.

University of Pennsylvania.

NORMAL BASES OF CYCLIC FIELDS OF PRIME-POWER DEGREE

By Sam Perlis

If Z is a normal field of degree n over F with automorphism group $G = (I, S_2, \cdots, S_n)$, a normal basis of Z/F is any basis of the form $(u, u^{S_2}, \cdots, u^{S_n})$. That such a basis always exists is well known (see [2; 32, footnote], [3], [4; bibliography], [5; bibliography]). We consider fields Z which are cyclic of degree $n = p^e$ over F, p being any prime, and obtain necessary and sufficient conditions for a quantity u of Z to generate a normal basis of Z/F. (When the conjugates of u form a basis of Z/F we shall say that u "generates" a normal basis of Z/F.) The structure of the extensions Z/F has been studied by A. A. Albert [1; IX] and his structure theorems are used here. The result is simplest in the case of characteristic p: the conjugates of a quantity u form a basis of Z/F if and only if the trace of u in Z/F is not zero. A particularly simple choice of u is given for this case. In the final section earlier results are applied to cyclotomic fields Z/F such that F contains a primitive p-th root of unity.

1. **Some lemmas.** We first consider cyclic fields of arbitrary finite degree over a field F.

Lemma 1. *Let Z be a cyclic field of degree n over F with generating automorphism S and a normal basis generated by a quantity v. Then a quantity $u = c_0 v + c_1 v^S + \cdots + c_{n-1} v^{S^{n-1}}$ of Z (c_i in F) generates a normal basis of Z/F if and only if the polynomials*

$$f(\lambda) = c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1}, \qquad g(\lambda) = \lambda^n - 1 \tag{1}$$

are relatively prime.

We must examine the matrix T expressing the conjugates of u in terms of the basis $(v, \cdots, v^{S^{n-1}})$, and find the conditions under which it is non-singular. This matrix is

$$T = \begin{Vmatrix} c_0 & c_1 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & \cdots & c_{n-2} \\ \cdots & \cdots & \cdots & \cdots \\ c_1 & c_2 & \cdots & c_0 \end{Vmatrix}. \tag{2}$$

If a matrix A is defined to be the special case of T obtained by setting $c_1 = 1$, and $c_i = 0$ for $i \neq 1$, we have

$$T = c_0 + c_1 A + \cdots + c_{n-1} A^{n-1} = f(A),$$

Received January 7, 1942; presented to the American Mathematical Society December 29, 1941.

where $f(\lambda)$ is the polynomial in (1). The polynomial $g(\lambda)$, furthermore, is both the characteristic and the minimum function of A. If $f(\lambda)$ and $g(\lambda)$ are relatively prime, there are polynomials $a(\lambda)$ and $b(\lambda)$ in the polynomial domain $F[\lambda]$ such that $a(\lambda)f(\lambda) + b(\lambda)g(\lambda) = 1$. Thus, since $g(A) = 0$, we find that $a(A)f(A) = 1$ and $T = f(A)$ is non-singular.

Conversely, suppose that $d = d(\lambda)$ of degree ≥ 1 is the g.c.d. of $f(\lambda) = f_1 d$ and $g(\lambda) = g_1 d$. Then $f(\lambda)g_1(\lambda) - g(\lambda)f_1(\lambda) = 0, f(A)g_1(A) - g(A)f_1(A) = f(A)g_1(A) = 0$, though $g_1(A) \neq 0$ since $g_1(\lambda)$ has degree less than n. The matrix $T = f(A)$ is thus a divisor of zero and cannot be non-singular. This completes the proof.

It is convenient to include in this section the proofs of two special results to be used later.

LEMMA 2. *Let F be a field of characteristic p, and $s_k = 1^k + 2^k + \cdots + (p-1)^k$ in F. Then $s_k = 0$ for $k = 1, 2, \cdots, p-2$ and $s_{p-1} = -1$.*

LEMMA 3. *Let p be a prime, and F be a field containing a primitive p-th root of unity, ζ, and $s_k = \sum_{i=0}^{p-1} (\zeta^i)^k$. Then $s_k = 0$ for $k = 1, 2, \cdots, p-1$ and $s_p = p$.*

Both results may be proved simultaneously by use of the Newton identities [1; 151, exercise 5] which assert that, if s_k is the sum of the k-th powers of the roots of an equation $\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0$, then $s_k + c_1 s_{k-1} + \cdots + c_{k-1}s_1 + kc_k = 0$ for $k = 1, 2, \cdots, n$. For the first result above, the quantities $1, 2, \cdots, p-1$ are the roots of the equation

$$\lambda^{p-1} - 1 = 0, \qquad n = p-1, \qquad c_1 = \cdots = c_{p-2} = 0, \qquad c_{p-1} = -1.$$

For the second result, the quantities $1, \zeta, \cdots, \zeta^{p-1}$ are the roots of

$$\lambda^p - 1 = 0, \qquad n = p, \qquad c_1 = \cdots = c_{p-1} = 0, \qquad c_p = -1.$$

Next we have a result holding for arbitrary normal fields of finite degree over F.

LEMMA 4. *Let N be a normal extension of degree $n = mq$ over F with a normal subfield L of degree m over F. Then, if u generates a normal basis of N/F, $v = T_{N/L}(u)$ generates a normal basis of L/F.*

If G is the group of N/F with normal subgroup H belonging to L, then $G = H + A_2H + \cdots + A_mH$, $HA_i = A_iH$ for every i, and the A_i together with $A_1 = I$ induce all the distinct automorphisms of L/F. Every v^{A_i} is in L and, in fact, $v^{A_i} = T_{N/L}(u^{A_i}) = u^{A_i} + u^{A_iH_1} + \cdots + u^{A_iH_q}$, where the H_j comprise H. It follows that the linear independence over F of the n quantities $u^{A_iH_j}$ implies that of the v^{A_i}.

2. *F* **has characteristic** *p.* In this section we consider cyclic p-fields, that is, cyclic extensions Z/F of degree p^e and characteristic p, and provide in Theorem 1 a determination of all normal bases of such extensions. The proof is a simple

application of the basic Lemma 1. The application of the latter result, however, requires a theorem insuring the existence of a normal basis, and the known proofs (see bibliography) of the general existence theorem, insofar as they require representation theory and linear algebras, are not as elementary as might be desired. In order to free our work of any dependence on the general existence proofs we shall state two further results yielding an explicit construction of an especially simple normal basis (see Theorem 3) and, although these results are immediate corollaries of Theorem 1, we shall prove them by methods which are elementary and completely independent of the earlier results of the paper.

THEOREM 1. *Let Z be a cyclic field of degree $n = p^e$ over F of characteristic p, and u be a quantity of Z. Then u generates a normal basis of Z/F if and only if the trace of u in Z/F is not zero.*

If u does generate a normal basis of Z/F, the trace of u is surely not zero, for otherwise we should have a conflict with the fact that the quantities of a basis are linearly independent. Conversely, suppose that this trace is not zero. In the notation of Lemma 1, $T_{Z/F}(u) = (c_0 + c_1 + \cdots + c_{n-1})T_{Z/F}(v)$ so that $c_0 + \cdots + c_{n-1} \neq 0$, that is, $f(1) \neq 0$. But in the present case $n = p^e$, $g(\lambda) = \lambda^n - 1 = (\lambda - 1)^n$ so that $\lambda = 1$ is the only root of $g(\lambda)$, and $f(\lambda)$ is prime to $g(\lambda)$. Thus it follows from Lemma 1 that u generates a normal basis of Z/F.

Before continuing with Theorems 2 and 3 we must recall known results [1; IX] about the structure of cyclic p-fields and establish notations. We have

$$Z = Z_e > Z_{e-1} > \cdots > Z_0 = F, \tag{3}$$

where each Z_i is cyclic of degree p^i over F and

$$Z_i = Z_{i-1}(\xi_i) = F(\xi_i), \qquad \xi_i^p = \xi_i + \alpha_i, \qquad \xi_i^S = \xi_i + \beta_i \tag{4}$$

with α_i and β_i in Z_{i-1}. Note that the generating automorphism S of Z/F induces a generating automorphism in each Z_i/F. We shall often write, for simplicity, Y for Z_{e-1}, ξ for ξ_e, α for α_e, β for β_e, and m for p^{e-1}.

THEOREM 2. *Let v generate an arbitrary normal basis of Y/F. Then $u = v\xi^{p-1}$ generates a normal basis of Z/F.*

We know that a basis of Z/F is given by the products

$$v^{S^k}\xi^j \qquad (k = 0, 1, \cdots, m-1; j = 0, 1, \cdots, p-1). \tag{5}$$

The proposed basis consists of $n = p^e$ quantities u^{S^i} which span a linear subspace L over F, and we shall show that $L = Z$ by showing that L contains every one of the quantities (5). We now require the following known consequence of the Newton identities. (See [1; 198f]. It is interesting to note that this lemma yields the following equivalent statement of Theorem 1: a quantity δ generates a normal basis of Z/F if and only if δ_{p-1} generates such a basis for Y/F.)

LEMMA 5. *Let* $\delta = \delta_0 + \delta_1\xi + \cdots + \delta_{p-1}\xi^{p-1}$, δ_i *in* Y. *Then* $T_{Z/Y}(\delta) = -\delta_{p-1}$.

Now L contains

$$u + u^{S^m} + \cdots + u^{S^{m(p-1)}} = vT_{Z/Y}(\xi^{p-1}) = -v. \tag{6}$$

For every quantity $\mu = \sum a_i u^{S^i}$ in L, every $\mu^{S^k} = \sum a_i u^{S^{i+k}}$ is also in L. It follows that L contains $v, v^S, \cdots, v^{S^{m-1}}$ so that L contains Y. We shall assume that L contains $v\xi^i$ for $i = 0, 1, \cdots, j-1 < p-1$ and shall prove that it contains $v\xi^j$; since $v\xi^{p-1}$ is in L we need consider only the case $j < p-1$. Let γ be the linear combination of the quantities

$$u = v\xi^{p-1}, \qquad u^{S^m} = v(\xi+1)^{p-1}, \qquad \cdots, \qquad u^{S^{m(p-1)}} = v(\xi+p-1)^{p-1}$$

with coefficients $1, 1 + \cdots + 1 = j+1, 2^0 + 2 + 2^2 + \cdots + 2^j, 3^0 + 3 + 3^2 + \cdots + 3^j, \cdots, 1 + (p-1) + \cdots + (p-1)^j$, respectively. Then the coefficient of $v\xi^{p-1}$ in γ is the sum of the coefficients just listed, that is, in the notation of Lemma 2, $p + s_1 + s_2 + \cdots + s_j = 0$. The coefficient of $v\xi^{p-1-k}$ for $k = 1, \cdots, p-1$ is

$$\begin{aligned}(-1)^k\{&[j+1]\cdot 1^k + [1 + 2 + \cdots + 2^j]\cdot 2^k + \cdots \\ &+ [1 + (p-1) + \cdots + (p-1)^j]\cdot(p-1)^k\}.\end{aligned} \tag{7}$$

The reason for this fact is that the coefficient of ξ^{p-1-k} in the expansion of $(\xi + r)^{p-1}$ is $r^k(p-1)(p-2)\cdots(p-k)/k! = r^k(-1)^k$ since $p = 0$. In the notation of Lemma 2 the expression (7) may be written as

$$(-1)^k(s_k + s_{k+1} + \cdots + s_{k+j}) \tag{8}$$

and this expression is equal to 0 whenever $k + j$ is less than $p-1$, $p-1-k > j$, and equal to $(-1)^{k+1}$ when $k + j = p-1$, $p-1-k = j$. This means that the coefficient of $v\xi^j$ in the expression γ is ± 1 and the coefficient of v times a higher power of ξ is 0. Then L contains $\pm\gamma = v(\xi^i + a_1\xi^{i-1} + \cdots + a_i)$, a_i in F, and by induction hypothesis L contains every $a_i v\xi^{i-i}$ $(i = 1, \cdots, j)$ so that these terms may be subtracted from $\pm\gamma$ leaving $v\xi^i$ in L. This completes the induction and proves that every $v\xi^i$ is in L $(i = 0, 1, \cdots, p-1)$.

Since $v\xi$ is in L it follows that every $v^{S^k}\xi^{S^k}$ is in L. But

$$v^{S^k}\xi^{S^k} = v^{S^k}(\xi + \beta + \beta^S + \cdots + \beta^{S^{k-1}}), \tag{9}$$

and every term $v^{S^k}\beta^{S^i}$ is in $Y \leq L$ and may be removed from the expression (9). This proves that $v^{S^k}\xi$ is in L for every k. We shall assume $v^{S^k}\xi^i$ in L for $i = 0, 1, \cdots, j-1$ and every k, and shall prove $v^{S^k}\xi^j$ in L. Since $v\xi^j$ is in L, so is

$$v^{S^k}(\xi^{S^k})^j = v^{S^k}(\xi + \beta + \cdots + \beta^{S^{k-1}})^j. \tag{10}$$

If we expand the expression in parentheses according to powers of ξ, we may write (10) in the form

$$v^{S^k}\xi^j + y_1\xi^{j-1} + y_2\xi^{j-2} + \cdots + y_j\,, \qquad y_i \text{ in } Y. \tag{11}$$

Each y_i is a linear combination of the conjugates of v, and every quantity $av^{S^k}\xi^f$ is in L for $f < j$ and a in F so that by subtracting such quantities from (11) we find that $v^{S^k}\xi^j$ is in L. We have proved that L contains every quantity of (5) so that $L = Z$.

THEOREM 3. *Let Z be cyclic of degree p^e over F of characteristic p with generation* (3), (4). *Then $u = (\xi_1\xi_2 \cdots \xi_e)^{p-1}$ generates a normal basis of Z over F.*

This result may be established as a consequence of Theorem 2 by a simple induction on e, provided we have proved it for $e = 1$. Hence, we now consider the case $e = 1$ and wish to prove that ξ^{p-1} generates a normal basis of $Z = Z_1$ over F. We need only express the quantities $\xi^{p-1}, (\xi+1)^{p-1}, \cdots, (\xi+p-1)^{p-1}$ in terms of $\xi^{p-1}, \xi^{p-2}, \cdots, \xi, 1$ and show the matrix of coefficients to be non-singular. This matrix M will be shown to be

$$\left\| \begin{array}{cccccc} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & -1 & 1 & \cdots & -1 & 1 \\ 1 & -2 & 2^2 & \cdots & -2^{p-2} & 2^{p-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & -(p-1) & (p-1)^2 & \cdots & -(p-1)^{p-2} & (p-1)^{p-1} \end{array} \right\|. \tag{12}$$

The correctness of the first row is obvious. The $(k+1)$-th row consists of the coefficients of $(\xi^{S^k})^{p-1} = (\xi+k)^{p-1} = \xi^{p-1} + (p-1)k\xi^{p-2} + \cdots + k^{p-1}$ which are of the form $k^t(p-1)(p-2)\cdots(p-t)/t! = (-1)^t k^t$. This establishes (12). Since the $(k+1)$-th row of (12) clearly is the vector $1, p-k, (p-k)^2, \cdots, (p-k)^{p-1}$ for $k = 0, 1, \cdots, p-1$, M is the Vandermonde matrix determined by the distinct quantities $1, 2, \cdots, p$. Thus, M is non-singular.

3. **F contains a primitive p-th root of unity.** We now consider cyclic fields Z of degree $n = p^e$ over F of characteristic not p. If the roots of the equation $\lambda^p - 1 = 0$ are not in F, this equation defines a field $C = F(\zeta_0)$, where ζ_0 is a primitive p-th root of unity and C/F has degree $\nu < p$. The composite $Z(\zeta_0)$ is the direct product $Z \times C$ over F and thus a quantity u of Z generates a normal basis of Z/F if and only if u generates a normal basis of $Z(\zeta_0)$ over $F(\zeta_0)$. We shall, therefore, study only the case in which F contains ζ_0.

For such fields F it is known [1; IX] that the extension Z/F has a generation of the following form:

$$Z = Z_e > Z_{e-1} > \cdots > Z_0 = F, \tag{13}$$

$$Z_i = Z_{i-1}(\xi_i), \qquad \xi_i^p = \alpha_i \text{ in } Z_{i-1}, \qquad \xi_i^S = \beta_i \xi_i, \tag{14}$$

$$\beta_i \text{ in } Z_{i-1}, \qquad N_{Z_{i-1}/F}(\beta_i) = \zeta_0, \tag{15}$$

for $i = 1, 2, \cdots, e$. Here S is a generating automorphism of Z/F. Especially important is the fact that when Z_i is given, a generation (14) may be found for each choice of β_i satisfying (15).

Further normalizations in the generation of Z above may be obtained by considering relations between the β_i and the possibility of choosing some of them in F. Let g be the largest integer such that the equation $\lambda^{p^{g-1}} = \zeta_0$ has a solution $\lambda = \zeta_{g-1}$ in F and define

$$\gamma_{g-i} = (\zeta_{g-1})^{p^i} = \zeta_{g-i-1} \qquad (i = 0, 1, \cdots, g-1). \tag{16}$$

If $g < e = f + g$, we note that, since

$$N_{Z_{e-1}/F}(\beta_e) = \zeta_0 = (\zeta_{g-1})^{p^{g-1}},$$

there must be [2; 97, Theorem 6] a quantity γ_e in Z_f such that

$$N_{Z_f/F}(\gamma_e) = \zeta_{g-1}, \qquad N_{Z_{e-1}/F}(\gamma_e) = \zeta_0, \tag{17}$$

and now we define

$$\gamma_{e-i} = N_{Z_f/Z_{f-i}}(\gamma_e) = N_{Z_{f-i+1}/Z_{f-i}}(\gamma_{e-i+1}) \qquad (i = 0, \cdots, f-1). \tag{18}$$

For every $j = 1, 2, \cdots, e$, the quantity γ_j is defined, is in Z_{j-1}, and satisfies the condition $N_{Z_{j-1}/F}(\gamma_j) = \zeta_0$. It follows that we may take every $\beta_j = \gamma_j$. We shall do so hereafter and shall refer to this fact by saying that Z/F has a "normalized" generation. In case F contains all of the n-th roots of unity, the normalized generation is the usual, simple one: $\xi^n = a$ in F, $\xi^S = \zeta\xi$, where $\zeta = \zeta_{e-1}$ is a primitive n-th root of unity.

We now obtain a lemma which is the analogue of Lemma 5 and in strong contrast with it. When viewed in the light of Lemma 4, this contrast shows that the present case must lead to results very different from those for the case of characteristic p. As in the previous section, we shall use the notations $Y = Z_{e-1}$, $\xi = \xi_e$, $\beta = \beta_e$, and $m = p^{e-1}$.

LEMMA 6. *Let* $\delta = \delta_0 + \delta_1\xi + \cdots + \delta_{p-1}\xi^{p-1}$, δ_i *in* Y. *Then* $T_{Z/Y}(\delta) = T_{Z/Y}(\delta_0) = p\delta_0$.

For, $T = S^m$ is a generating automorphism of Z/Y,

$$\xi^T = \beta\beta^S \cdots \beta^{S^{m-1}}\xi, \qquad \xi^{T^j} = \zeta_0^j\xi, \qquad (\xi^i)^{T^j} = (\zeta_0^j)^i\xi^i,$$

and

$$T_{Z/Y}(\delta) = p\delta_0 + \delta_1\xi s_1 + \cdots + \delta_{p-1}\xi^{p-1}s_{p-1}, \qquad s_k = \sum_{j=0}^{p-1} (\zeta_0^j)^k$$

for $k = 1, 2, \cdots, p-1$. Now s_k is the sum of the k-th powers of the roots of the equation $\lambda^p - 1 = 0$, and Lemma 3 informs us that $s_k = 0$ for $k < p$, as desired.

THEOREM 4. *Let Z be cyclic of degree $n = p^e$ over F containing a primitive p-th root of unity, p a prime, so that Z has the normalized generation described above. Let h be any fixed non-negative integer on the range $e-1, e-2, \cdots, e-g$, $s = p^h$, $r = p^{e-h} = n/s$, and let*

$$u = \delta_0 + \delta_1\xi + \cdots + \delta_{r-1}\xi^{r-1} \qquad (\delta_i \text{ in } Z_h) \tag{19}$$

be any quantity of Z. Then u generates a normal basis of Z/F if and only if $Z_k = L_k$ $(k = 0, 1, \cdots, r-1)$, where L_k is the linear set over F spanned by the quantities

$$\delta_k, \quad \delta_k^S\beta^k, \quad \delta_k^{S^2}(\beta\beta^S)^k, \quad \cdots, \quad \delta_k^{S^{t-1}}(\beta\cdots\beta^{S^{t-2}})^k. \tag{20}$$

(This theorem implies that the result of Theorem 1 is not valid for the present case. For, we may take $\delta_k = 0$ for $k > 0$ so that $L_k = 0 \neq Z_k$, but still retain, by Lemma 6, the property $T_{Z/F}(u) \not\equiv 0$.)

The condition $L_0 = Z_h$ means simply that δ_0 generates a normal basis of Z_h/F. The theorem provides g different sets of necessary and sufficient conditions (or e sets if $g \geq e$); for the choice $h = e-1$ the proof to be given is valid for an arbitrary generation (13), (14), (15).

Let $t = p^f$, where $f = e - g$ if $e \geq g$ and, otherwise, $f = 0$. Then $\beta\beta^S \cdots \beta^{S^{t-1}} = N_{Z/F}(\beta) = \zeta_{r-f-1}$ and $s = p^h = tp^{h-f}$ so that

$$\beta\beta^S \cdots \beta^{S^{s-1}} = (\zeta_{r-f-1})^{p^{h-f}} = \zeta_{r-h-1} = \rho \tag{21}$$

and (21) is in F and is a primitive r-th root of unity. The conjugates of u may then be displayed in r sets of s conjugates in the following way:

$$u = \delta_0 + \delta_1\xi + \cdots + \delta_{r-1}\xi^{r-1};$$

$$u^S = \delta_0^S + \delta_1^S\beta\xi + \cdots + \delta_{r-1}^S\beta^{r-1}\xi^{r-1};$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$u^{S^{s-1}} = \delta_0^{S^{s-1}} + \delta_1^{S^{s-1}}(\beta\beta^S\cdots\beta^{S^{s-2}})\xi + \cdots + \delta_{r-1}^{S^{s-1}}(\beta\cdots\beta^{S^{s-2}})^{r-1}\xi^{r-1};$$

$$u^{S^s} = \delta_0 + \delta_1\rho\xi + \cdots + \delta_{r-1}\rho^{r-1}\xi^{r-1};$$

$$u^{S^{s+1}} = \delta_0^S + \delta_1^S\rho\beta\xi + \cdots + \delta_{r-1}(\rho\beta)^{r-1}\xi^{r-1};$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$u^{S^{n-1}} = \delta_0^{S^{s-1}} + \delta_1^{S^{s-1}}(\rho^{r-1}\beta\beta^S\cdots\beta^{S^{s-2}})\xi + \cdots + \delta_{r-1}^{S^{s-1}}(\rho^{r-1}\beta\cdots\beta^{S^{s-2}})^{r-1}\xi^{r-1}.$$

The n quantities u^{S^i} span a linear subspace L over F of Z, all quantities of L having the form $\gamma_0 + \gamma_1\xi + \cdots + \gamma_{r-1}\xi^{r-1}$ with each γ_k in the space L_k spanned by the quantities (20). Since $1, \xi, \cdots, \xi^{r-1}$ form a basis of Z/Z_h and $L_k \leq Z_h$ for each k, it is necessary that every $L_k = Z_h$ if $L = Z$.

Conversely, suppose that every $L_k = Z_k$ and let $\gamma = a_0 u + a_1 u^S + \cdots + a_{n-1} u^{S^{n-1}} = 0$ with a_i in F. If we write $\gamma = \gamma_0 + \gamma_1 \xi + \cdots + \gamma_{r-1} \xi^{r-1}$ with the γ_k in Z_h, we must have each $\gamma_k = 0$. We find that γ_k is a linear combination of the quantities (20) with coefficients given by

$$\alpha_j = a_j + a_{j+s} \rho^k + a_{j+2s} (\rho^k)^2 + \cdots + a_{j+(r-1)s} (\rho^k)^{r-1}$$

for $j = 0, 1, \cdots, s-1$. The hypotheses $L_k = Z_k$ and $\gamma_k = 0$ mean that each $\alpha_j = 0$. Thus the s equations

$$f_j(\lambda) = a_j + a_{j+s} \lambda + \cdots + a_{j+(r-1)s} \lambda^{r-1} = 0 \tag{22}$$

of degree at most $r-1$ have the r distinct roots ρ^k ($k = 0, 1, \cdots, r-1$), whence we conclude that every $a_i = 0$ so that the u^{S^i} are linearly independent over F.

When F contains the n-th roots of unity, β is in F so that each L_k is spanned by $\delta_k, \cdots, \delta_k^{S^{n-1}}$. This yields the following result.

COROLLARY. *Let F contain the n-th roots of unity. Then u of* (19) *generates a normal basis of Z/F if and only if every δ_k generates a normal basis of Z_h/F.*

Since $g \geq e$ now, we may choose $h = 0$ and then the condition of the corollary becomes simply that every $\delta_k \neq 0$. This result may be generalized to the case of arbitrary degree. Let Z/F be cyclic of arbitrary finite degree q and let F contain a primitive q-th root of unity, so that $Z = F(\xi)$, $\xi^q = a$ in F. Then a quantity $u = \delta_0 + \delta_1 \xi + \cdots + \delta_{q-1} \xi^{q-1}$ (δ_k in F) generates a normal basis of Z/F if and only if every $\delta_k \neq 0$.

For proof one expresses the conjugates of u in terms of the powers of ξ and finds that the matrix of coefficients is a Vandermonde matrix, determined by the q-th roots of unity, whose columns have been multiplied by the respective coefficients δ_k of u. This yields the result.

4. **Cyclotomic fields.** The result of Theorem 4 may now be applied to cyclotomic fields. Let C_e be the root field over F of the equation

$$\lambda^{p^{e+1}} - 1 = 0 \tag{23}$$

so that $C_0 = F$ by hypothesis. Let us assume, now, that F contains no roots of (23) other than the p-th roots of unity. As in the previous section we let ζ_e denote a primitive root of (23) and define $\zeta_j = \zeta_e^{p^{e-j}}$ ($j = 0, 1, \cdots, e$) so that ζ_j is a primitive root of $\lambda^{p^{j+1}} = 1$.

Since the equation $\lambda^p - \zeta_0 = 0$ has no root in F, it is irreducible [1; 188, Theorems 21 and 22] over F and defines a cyclic field $C_1 = F(\zeta_1)$, $\zeta_1^p = \zeta_0$, $\zeta_1^S = \zeta_0 \zeta_1$. For $p > 2$ we shall prove by a simple application of the theorems of Albert the known fact that C_e/F is cyclic of degree p^e, and as a by-product shall obtain an explicit form for the generation (13), (14), (15) of C_e/F, $C_i = Z_i$.

THEOREM 5. *For $p > 2$, the cyclotomic extension C_e/F described above is cyclic of degree p^e and has generation* (13), (14), (15) *in which $Z_i = C_i$, $\xi_i = \zeta_i$ and $\alpha_i = \beta_i = \zeta_{i-1}$ for $i = 1, 2, \cdots, e$. Further, it has a normal basis generated by the quantity*

$$u = 1 + (\zeta_1 + \cdots + \zeta_1^{p-1}) + \cdots + (\zeta_e + \cdots + \zeta_e^{p-1}).$$

The first assertion was discussed above for the case $e = 1$ and we shall make an induction on e, assuming that $C_{e-1} = F(\zeta_{e-1})$ is cyclic of degree p^{e-1} and $\zeta_{e-1}^S = \zeta_{e-2}\zeta_{e-1}$. Since $\beta_{e-1} = \zeta_{e-2}$, we have

$$N_{C_{e-2}/F}(\zeta_{e-2}) = \zeta_0 . \tag{24}$$

Now ζ_{e-1} has the minimum equation $\lambda^p - \zeta_{e-2}$ over C_{e-2} so that the norm of ζ_{e-1} in C_{e-1}/C_{e-2} is ζ_{e-2} when p is odd. Combining this fact with (24) we obtain

$$N_{C_{e-1}/F}(\zeta_{e-1}) = \zeta_0 . \tag{25}$$

The existence in C_{e-1} of a quantity ζ_{e-1} satisfying (25) implies [1; IX, Theorems 11, 12] that C_{e-1} has overfields which are cyclic of degree p^e over F and for which $\beta_e = \zeta_{e-1}$. Since $\beta_e^p = \zeta_{e-2} = (\zeta_{e-1}^S)/\zeta_{e-1}$, one such overfield is defined by the equation $\lambda^p = \zeta_{e-1} = \alpha_{e-1}$. This field is C_e and the induction is complete.

To prove the second assertion in Theorem 5 we note that for $e = 1$ it is proved by the corollary to Theorem 4 so that we may use induction on e, assuming the result for $e - 1$. Using Theorem 4 with $h = e - 1$ and $r = p$, we write $u = \delta_0 + \zeta_e + \cdots + \zeta_e^{p-1}$ so that in (19) $\xi = \zeta_e$ and $\delta_k = 1$ for $k > 0$. The induction hypothesis means that δ_0 generates a normal basis of Z_{e-1}/F, and it remains only to show $L_k = Z_{e-1}$ $(k = 1, 2, \cdots, p - 1)$, noting that in (20) we now have $\delta_k = 1$ and $\beta = \zeta_{e-1}$. Since k is prime to p, $\omega = \zeta_{e-1}^k$ is also a primitive n-th root of unity; hence a basis of $C_{e-1} = F(\zeta_{e-1}) = F(\omega)$ is given by

$$(1, \omega, \omega^2, \cdots, \omega^{s-1}), \qquad s = p^{e-1}.$$

The quantities (20) now have the form

$$1, \quad \omega, \quad \omega\omega^S, \quad \cdots, \quad \omega\omega^S \cdots \omega^{S^{s-2}}, \tag{26}$$

and since these s quantities are all powers of ω with ω^s in F, it suffices to show that the exponents on these powers are incongruent modulo $s = p^{e-1}$. Noting that $\zeta_{e-1}^S = \zeta_{e-2}\zeta_{e-1} = (\zeta_{e-1})^{p+1}$, $(\zeta_{e-1}^k)^S = \omega^S = \omega^{p+1}$, we find these exponents to be

$$\begin{aligned} &0, \quad 1, \quad 1 + (p+1), \quad 1 + (p+1) + (p+1)^2, \quad \cdots, \\ &\qquad 1 + (p+1) + \cdots + (p+1)^{s-2}. \end{aligned} \tag{27}$$

First, we shall show that none of these quantities, except the first, is divisible by p^{e-1}, or, what is the same, none of the quantities (26) is in F except the first.

Suppose that q is the smallest positive integer such that $w = \omega\omega^S \cdots \omega^{S^{q-1}}$ is in F. Then $n = qt + c$, $0 \leq c < q$, and if $c > 0$ we have

$$\omega\omega^S \cdots \omega^{S^{n-1}} = 1 = ww^{S^q} \cdots w^{S^{q(t-1)}} w_0^{S^{qt}} = w^t w_0^{S^{qt}}$$

with

$$w_0 = \omega\omega^S \cdots \omega^{S^{c-1}} = (w^{-t})^{S^{n-qt}} = w^{-t} \qquad \text{in } F.$$

This contradiction with the definition of q implies that $c = 0$, q must be a power of p, and we need only show that none of the quantities

$$b_j = 1 + (p+1) + \cdots + (p+1)^{p^j - 1} \qquad (j = 1, 2, \cdots, e-2) \tag{28}$$

is divisible by p^{e-1}.

We shall make an induction on e, the first case to occur being $e = 3$. Then j must be 1,

$$\begin{aligned} b_1 &\equiv 1 + (p+1) + (2p+1) + \cdots + (p-1)p + 1 \\ &\equiv p + p \cdot p(p-1)/2 \equiv p \pmod{p^2} \end{aligned}$$

since p is odd. Thus $b_1 \not\equiv 0 \pmod{p^2}$ and the case $e = 3$ is completed. Now we assume that for every $j \leq e-3$, $b_j \not\equiv 0 \pmod{p^{e-2}}$ and wish to prove

$$b_j \not\equiv 0 \pmod{p^{e-1}} \qquad (j = 0, 1, \cdots, e-2).$$

For $j < e-2$, this is obvious. Now

$$b_j = b_{j-1}t_j = b_{j-1}[1 + (p+1)^{p^{j-1}} + (p+1)^{2p^{j-1}} + \cdots + (p+1)^{(p-1)p^{j-1}}$$

so that $b_{e-2} = b_{e-3}t_{e-2}$. Since the highest possible power of p dividing b_{e-3} is p^{e-3}, it suffices to show that t_{e-2} does not have a factor p^2. But

$$t_{e-2} \equiv 1 + 1 + \cdots + 1 = p \pmod{p^2},$$

$$t_{e-2} \not\equiv 0 \pmod{p^2},$$

and the induction is completed.

We know, now, that none of the non-zero quantities in (27) is divisible by p^{e-1}. If two of the quantities in (27) were congruent modulo p^{e-1}, we should immediately obtain a contradiction with the fact just proved. Theorem 5 is thus established.

Suppose that we modify the quantity u of the theorem by giving each of its terms ζ_i^j and 1 an arbitrary non-zero coefficient in F. Then it is easy to see that Theorem 5 is valid for this generalized u and that the induction proof just given still holds. For, the quantities (20) all acquire a factor in F and this does not affect the linear independence over F of these quantities.

COROLLARY. *Every quantity*

$$u = \alpha_{01} + (\alpha_{11}\zeta_1 + \cdots + \alpha_{1,p-1}\zeta_1^{p-1}) + \cdots + (\alpha_{s1}\zeta_s + \cdots + \alpha_{s,p-1}\zeta_s^{p-1})$$

with all coefficients α_{ij} *in* F *and* $\neq 0$ *generates a normal basis of the extension* Z/F *of Theorem* 5.

BIBLIOGRAPHY

1. A. A. ALBERT, *Modern Higher Algebra*, Chicago, 1937.
2. A. A. ALBERT, *Structure of Algebras*, American Mathematical Society Colloquium Publications, vol. XXIV, New York, 1939.
3. M. DEURING, *Galoissche Theorie und Darstellungstheorie*, Mathematische Annalen, vol. 107(1932), pp. 140–144.
4. T. NAKAYAMA, *Normal basis for a quasi-field*, Proceedings of the Imperial Academy of Tokyo, vol. 16(1940), pp. 532–536.
5. R. STAUFFER, *The construction of a normal basis in a separable normal extension field*, American Journal of Mathematics, vol. 58(1936), pp. 585–597.

LOCKHEED AIRCRAFT CORPORATION, BURBANK, CALIFORNIA.

COMPLETELY MONOTONE FUNCTIONS IN PARTIALLY ORDERED SPACES

By S. Bochner

1. The theorem on completely monotone functions.

Theorem (Hausdorff-Bernstein-Widder [4]). *If $T(\alpha)$ is defined in $0 < \alpha < \infty$ and if for each α and h $(0 < h < \infty)$ the relations*

$$(1) \qquad (-1)^n \Delta_h^n T(\alpha) \geq 0, \qquad n = 0, 1, 2, \cdots$$

hold, where $\Delta_h^0 T(\alpha) \equiv T(\alpha)$, $\Delta_h^1 T(\alpha) = T(\alpha + h) - T(\alpha)$ and $\Delta_h^{n+1} T = \Delta_h^1(\Delta_h^n T)$, then $T(\alpha)$ can be represented in the form

$$(2) \qquad T(\alpha) = \int_0^\infty e^{-\alpha t}\, dE(t)$$

and vice versa, where $E(t)$ is a function in $0 \leq t < \infty$ for which $E(0) = 0$, and $\Delta E(t) \geq 0$.

Also, the limits $E(+0)$ and $E(t \pm 0)$, $0 < t < \infty$, are uniquely determined.

It is the purpose of the present note to point out that the theorem remains true if the values of $T(\alpha)$ and $E(t)$ instead of being numbers are elements of a suitable vector space S and that the space S may be as general as the wording of the theorem will allow.

In order to be able to state assumption (1), it is sufficient to require that S be a commutative group of addition and that it be partially ordered by a relation $T \geq 0$ with the properties:

(i) $T \geq T$,
(ii) $T \geq U$, $U \geq T$ *imply* $T = U$,
(iii) $T \geq U$, $U \geq V$ *imply* $T \geq V$,
(iv) $T \geq U$ *implies* $T + V \geq U + V$ *for any* V.

It is customary to add the following property:

(v) *Given T and U there exists an element V such that $V \geq T$, $V \geq U$,*

but we emphasize that this property will not be needed.

Let us leave (1). However, as soon as we envisage the prospective relation (2), we are faced with the necessity of defining a Stieltjes integral of the type

$$\int_a^b \varphi(t)\, dE(t),$$

Received February 5, 1942.

where $\varphi(t)$ is a number and $E(t)$ is an element of S for each t. The definition of such an integral necessitates the existence within S of sums $\sum_{m=1}^{n} a_m T_m$ for arbitrary real coefficients a_m and of limits for increasing (and, dually, decreasing) sequences of such sums. We therefore add the following properties.

(vi) *S is a vector space with real coefficients, and $T \geq 0$, $a \geq 0$ imply $aT \geq 0$.*

(vii) *Every monotone increasing sequence which is bounded from above has a least upper bound which will be denoted by "sup". That is, if $T_n \leq T_{n+1}$, $n = 1, 2, \cdots$, and $T_n \leq U$ for some U, then there exists an element $T_0 = \sup_n T_n$ such that $T_n \leq T_0$, and $T_0 \leq U$ if U is any element for which $T_n \leq U$.*

We will see in §2 that these properties are indeed sufficient to define the Stieltjes integral adequately and that our theorem holds. As pointed out before, no other lattice property will be required, not even property (v). If T is a symmetric operator in Hilbert space, we put $T \geq 0$ whenever T is positive definite. With this partial ordering, we obtain spaces S which are not lattices. They are the space of all bounded Hermitian operators and suitable subspaces of that space. Certain spaces of unbounded symmetric operators are also admissible, as for instance, to mention a trite extension of bounded operators, the space of all symmetric operators which are represented by all symmetric forms of type $\sum a_{m,n} x_m x_n$.

A peculiar situation arises if S is also a ring with a unit element 1 for which (a) $1 > 0$ and (b) $T \geq 0$ and $U \geq 0$ imply $TU \geq 0$. Under these circumstances we may consider a semi-group $T(\alpha)$ in $0 < \alpha < \infty$. It is a function with the property

$$T(\alpha)\, T(\beta) = T(\alpha + \beta). \tag{3}$$

If in addition

$$0 < T(\alpha) \leq 1, \tag{4}$$

assumption (1) will be automatically fulfilled since

$$-\Delta_h^1\, T(\alpha) = T(\alpha) - T(\alpha + h) = T(\alpha)\,(1 - T(h)) \geq 0$$

and, in general,

$$(-1)^n \Delta_h^n\, T(\alpha) = T(\alpha)\,(1 - T(h))^n \geq 0.$$

Hence our theorem applies and we have a representation (2). Now, on the basis of (3), if $\varphi(t)$ and $\psi(t)$ are each a *finite* sum of the form $\sum_\nu a_\nu e^{-\alpha_\nu t}$ (a_ν real, $\alpha_\nu > 0$), we immediately obtain relation

$$\int_0^\infty \varphi(t)\, dE(t) \cdot \int_0^\infty \psi(t)\, dE(t) = \int_0^\infty \varphi(t)\, \psi(t)\, dE(t). \tag{5}$$

Furthermore, if in S the product TU is a "continuous" function of its factors, we may pass to limits in relation (5) and establish its validity for any continuous

functions $\varphi(t)$, $\psi(t)$ in $0 \leq t < \infty$ which vanish outside a finite interval. Also, on the basis of (4) we obtain the relation

$$0 \leq E(t) \leq 1 \qquad (0 \leq t < \infty), \tag{6}$$

and relations (5) and (6) taken together express the familiar fact that $E(t)$ is a "resolution of a part of the identity".

We thus obtain a recent result of E. Hille [2] to the effect that for commutative bounded Hermitian operators $T(\alpha)$ the assumptions (3) and (4) imply a formula (2) with (5) and (6).

All our results can also be proved for the "discrete" case in which α and h assume only integer values $1, 2, 3, \cdots$. In this case, a semi-group $T(n)$ can be obtained from any element T by putting $T(n) = T^n$. Hence, for $0 \leq T \leq 1$, we obtain

$$T = \int_0^\infty e^{-t}\, dE(t) = \int_{+0}^1 \tau d_\tau[-E(\log \tfrac{1}{\tau})] \tag{7}$$

and if T is an Hermitian operator, this is its spectral formula and our derivation is in substance the method of F. Riesz [3].

2. **Proof of the theorem.** We will first list additional properties of S with brief proofs. They will be used frequently, sometimes without explicit reference.

(i) If $\{T_n\}$ and $\{U_m\}$ are both increasing and each $T_n \geq$ some U_m and each $U_m \geq$ some T_n, then

$$\sup T_n = \sup U_m .$$

This holds, in particular, if $\{U_m\}$ is a subsequence of $\{T_n\}$. The proof follows readily from the definition of sup.

(ii) If $\{T_n\}$ and $\{U_n\}$ are both increasing, then

$$\sup (T_n + U_n) = \sup T_n + \sup U_n .$$

In fact, the relation $T_n + U_n \leq \sup T_n + \sup U_n$ immediately leads to $\sup (T_n + U_n) \leq \sup T_n + \sup U_n$. On the other hand, for $m \leq n$, $T_m + U_n \leq T_n + U_n \leq \sup (T_n + U_n)$; hence $U_n \leq \sup (T_n + U_n) - T_m$; hence $\sup T_m \leq \sup (T_n + U_n) - \sup U_n$; hence $\sup T_n + \sup U_n \leq \sup (T_n + U_n)$.

(iii) If $\{T_n\}$ is increasing and $\lambda > 0$, then

$$\sup (\lambda T_n) = \lambda \sup T_n .$$

In fact, $\lambda T_n \leq \lambda \sup T_n$; hence $\sup (\lambda T_n) \leq \lambda \sup T_n$; hence also $\lambda \sup T_n = \lambda \sup (\lambda^{-1}\lambda T_n) \leq \lambda\lambda^{-1} \sup (\lambda T_n) = \sup (\lambda T_n)$.

(iv) If $T < 0$ and $\epsilon_n \searrow 0$ (ϵ_n is a real number) then

$$\sup (\epsilon_n T) = 0.$$

In fact, there is a subsequence $\{\epsilon_{n_k}\}$ of $\{\epsilon_n\}$ for which

$$\epsilon_{n_{k+1}} < 2\epsilon_{n_{k+1}} \leq \epsilon_{n_k} .$$

Hence, by (i) and (iii), $\sup(\epsilon_n T) \leq 2 \sup(\epsilon_n T) \leq \sup(\epsilon_n T)$, or $\sup(\epsilon_n T) = 2 \sup(\epsilon_n T)$, or $\sup(\epsilon_n T) = 0$.

The key to our argument will be the following simple lemma.

LEMMA 1. *If $T_n \leq T_{n+1} \leq U_{n+1} \leq U_n$, $n = 1, 2, \cdots$, and if there exist an element $V > 0$ and a sequence $\epsilon_n \searrow 0$ such that*

$$U_n - T_n \leq \epsilon_n V, \tag{8}$$

then

$$\sup T_n = \inf U_n . \tag{9}$$

In fact, the inequality $\sup T_n \leq \inf U_n$ is obvious. Now, (8) can be written in the form

$$-\epsilon_n V \leq T_n + (-U_n)$$

and hence, by (ii) and (iv),

$$0 = \sup(-\epsilon_n V) \leq \sup T_n + \sup(-U_n) = \sup T_n - \inf U_n ,$$

that is, $\inf U_n \leq \sup T_n$.

DEFINITIONS. The function space F consists of all real continuous functions $f(t)$ in $0 \leq t < \infty$ for which

$$|f(t)| \leq ae^{-\alpha t}, \qquad a > 0, \qquad \alpha > 0, \qquad 0 \leq t < \infty, \tag{10}$$

where the constants a and α depend on $f(t)$. An element $f(t)$ of F is called a special function if it vanishes outside some interval $0 \leq t \leq t_0$, t_0 depending on $f(t)$, and it is called pseudo-polynomial if it is a finite sum

$$p(t) = \sum_{\nu=1}^{\infty} c_\nu e^{-\alpha_\nu t}, \tag{11}$$

where c_ν is real, and $\alpha_\nu > 0$ and *rational*. The set of special functions will be denoted by F_0, and the set of pseudo-polynomials by P.

A distributive operation $A(f)$ from a set F^0 in F to S is called positive if $f(t) \geq 0$ implies $A(f) \geq 0$.

LEMMA 2. *The set P is dense in the space F in the following sense. Corresponding to any element $f \in F$ there exist sequences $\{p_n\}$, $\{q_n\}$ in P, an exponent $\alpha > 0$, and a sequence $\epsilon_n \searrow 0$ such that*

$$\begin{aligned} &p_n(t) \leq p_{n+1}(t) \leq f(t) \leq q_{n+1}(t) \leq q_n(t), \\ &q_n(t) - p_n(t) \leq \epsilon_n e^{-\alpha t} \qquad (n = 1, 2, \cdots ; 0 \leq t < \infty). \end{aligned} \tag{12}$$

In fact, if $|f(t)| \leq ae^{-\beta t}$, then for $0 < \alpha < \beta$, $e^{\alpha t} f(t)$ is uniformly continuous in $0 \leq t < \infty$. By Weierstrass' approximation theorem, corresponding to any sequence $\eta_n \searrow 0$ there exists a sequence of ordinary polynomial $\varphi_n(t)$ in e^{-t}, such that $|e^{\alpha t} f(t) - \varphi_n(t)| \leq \eta_n$, that is,

$$|f(t) - e^{-\alpha t}\varphi_n(t)| \leq \eta_n e^{-\alpha t}. \tag{13}$$

Now, if $\sum \eta_n$ is convergent and if we put $\rho_n = \sum_{m=n}^{\infty} (\eta_m + \eta_{m+1})$, then the quantities

$$p_n(t) = e^{-\alpha t}\varphi_n(t) - \rho_n e^{-\alpha t}, \qquad q_n(t) = e^{-\alpha t}\varphi_n(t) + \rho_n e^{-\alpha t}, \qquad \epsilon_n = 2\rho_n$$

satisfy (12).

LEMMA 3. *Any positive operation Af from P to S can be continued to a positive operation from F to S. The continuation is unique.*

In fact, if we put $T_n = A(p_n)$, $U_n = A(q_n)$, $V = A(e^{-\alpha t})$, then, by Lemma 1, (12) implies

$$\sup T_n = \inf U_n. \tag{14}$$

Given $f \in F$, it is not hard to verify that the common value of both sides in (14) is independent of the approximating sequences (12), and it is also not hard to verify that (14) defines a distributive operation on F. If $f(t) \geq 0$, each $q_n(t) \geq 0$, and hence $A(f) \geq 0$. The proof for the uniqueness of the continuation is implied in the proof for its existence.

LEMMA 4. *If $p(t)$ is an element of P, and $p(t) > 0$ in $0 \leq t < \infty$, then $p(t)$ is a finite linear combination with positive coefficients of functions $e^{-\beta t}(1 - e^{-\gamma t})^n$, where β and γ are positive rational numbers and $n = 0, 1, 2, \cdots$.*

In fact, there is a positive rational exponent α such that $e^{\alpha t}\, p(t)$ is an ordinary polynomial of $e^{-\gamma t}$, where γ is a suitable rational exponent (reciprocal of an integer). Hence, by a fundamental theorem of Hausdorff [1], $e^{\alpha t}p(t)$ is a finite linear combination with positive coefficients of expressions $e^{-m\gamma t}(1 - e^{-\gamma t})^n$ $(m, n = 0, 1, 2, \cdots)$. This proves the lemma.

Now consider the function $T(\alpha)$ of our theorem and for every rational α, put $A(e^{-\alpha t}) = T(\alpha)$. This defines a distributive operation from P to S. By (1), $A(e^{-\alpha t}(1 - e^{-\beta t})^n) \geq 0$, and, by Lemma 4, $A(p) \geq 0$ if $p(t) > 0$. If $p(t) \geq 0$, we take any rational $\alpha > 0$ and consider the polynomial $p(t) + \epsilon e^{-\alpha t}$ for $\epsilon > 0$. Since it > 0, we have $A(p) \geq -\epsilon T(\alpha)$, and by property (iv), this leads to $A(p) \geq 0$. Hence, $A(p)$ is a positive operation, and, by Lemma 3, it can be continued onto all of F.

If $\alpha > 0$ is irrational, there exist positive rational numbers β_n, γ_n, δ and $\epsilon_n \searrow 0$ such that $\beta_n \leq \beta_{n+1} \leq \alpha \leq \gamma_{n+1} \leq \gamma_n$ and $e^{-\beta_n t} - e^{-\gamma_n t} \leq \epsilon_n e^{-\delta t}$. Hence,

applying Lemma 1 to the quantities $T_n = T(\gamma_n) = A(e^{-\gamma_n t})$, $U_n = T(\beta_n) = A(e^{-\beta_n t})$, $V = A(e^{-\delta t})$, we readily conclude

$$\sup_\epsilon T(\alpha - \epsilon) = A(e^{-\alpha t}) = \inf_\epsilon T(\alpha + \epsilon).$$

Thus $T(\alpha)$ is continuous, and $T(\alpha) = A(e^{-\alpha t})$ for all $\alpha > 0$, rational or irrational.

If $\rho > 0$ and $0 < \epsilon < \rho$, we consider the special function $\omega_{\rho\epsilon}(t)$ which is 1 in $0 \leq t \leq \rho - \epsilon$, 0 in $\rho \leq t < \infty$, and linear in $\rho - \epsilon \leq t \leq \rho$. If ϵ decreases, $\omega_{\rho\epsilon}(t)$ increases, and hence the value

$$\sup_\epsilon A(\omega_{\rho\epsilon}(t)) = E(\rho) \tag{15}$$

exists. This is the function $E(t)$ to appear in (2).

We pick a special function $f(t)$ and a number $a > 0$ such that $f(t) = 0$ for $t \geq a$. We consider any points

$$0 = t_0 < t_1 < t_2 < \cdots < t_n < t_{n+1} = a, \tag{16}$$

any numbers l_m, and any numbers $\eta_m > 0$ such that

$$f(t) \geq l_m + \eta_m, \qquad t_m \leq t \leq t_{m+1}, \qquad (m = 0, 1, \cdots, n). \tag{17}$$

If ϵ is smaller than the mesh of the partition (16), then the continuous function

$$g_\epsilon(t) = l_0\, \omega_{t_1\epsilon} + \sum_{m=1}^{n} l_m (\omega_{t_{m+1}\epsilon} - \omega_{t_m\epsilon})$$

has the constant value l_m in $(t_m, t_{m+1} - \epsilon)$, is linear in the connecting intervals $(t_{m+1} - \epsilon, t_{m+1})$, and is 0 for $t \geq a$. Therefore, for ϵ sufficiently small (depending on the η_m), $f(t) \geq g_\epsilon(t)$ in $0 \leq t < \infty$. Thus, $A(f) \geq A(g_\epsilon)$, that is,

$$A(f) \geq l_n\, A(\omega_{t_{n+1}\epsilon}) - \sum_{m=1}^{n} (l_m - l_{m+1}) A(\omega_{t_m\epsilon}). \tag{18}$$

Since $A(\omega_{t_m\epsilon})$ is increasing with ϵ decreasing, we may first let ϵ tend to 0 in the second term on the right side of (18). This done we may also let $\epsilon \searrow 0$ in the first term, and we thus obtain

$$A(f) \geq l_n\, E(t_{n+1}) - \sum_{m=1}^{n} (l_m - l_{m-1}) E(t_m), \tag{19}$$

that is,

$$A(f) \geq J(l_m, t_m), \tag{20}$$

where

$$J(l_m, t_m) = l_0\, E(t_1) + \sum_{m=1}^{n} l_m (E(t_{m+1}) - E(t_m)). \tag{21}$$

The sum (21) increases with each l_m. Hence, by applying sup successively we may let $\eta_m \to 0$, and finally define

$$l_m = \min_{t_m \leq t \leq t_{m+1}} f(t).$$

Similarly we have

$$A(f) \leq J(\lambda_m, t_m), \tag{22}$$

where

$$\lambda_m = \max_{t_m \leq t \leq t_{m+1}} f(t). \tag{23}$$

Now, if $E(t)$ is any increasing function in $0 \leq t \leq a$, with $E(0) = 0$, and $f(t)$ is a continuous real function, if we take a directed sequence of partitions π_k ($k = 1, 2, 3, \cdots$), each of the form (16), if we denote the corresponding lower and upper Darboux sums $J(l_m^k, t_m^k)$ and $J(\lambda_m^k, t_m^k)$ by T_k and U_k respectively, and if for the k-th partition we put $\epsilon_k = \max_m (\lambda_m^k - l_m^k)$, then

$$U_k - T_k \leq \epsilon_k E(a).$$

Hence, by Lemma 1,

$$\sup T_k = \inf U_k$$

and the common value may be denoted by

$$\int_0^a f(t)dE(t).$$

Also, by (20) and (22) we may now write

$$A(f) = \int_0^\infty f(t)dE(t), \tag{24}$$

where f is any special function and $\int_0^\infty$ indicates the common value of $\int_0^b$ for all $b \geq a$, where $f(t) = 0$ for $t \geq a$.

Finally, if $f(t) = e^{-\alpha t}$, for some $\alpha > 0$, we may pick a sequence of number $0 < a_1 < a_2 < \cdots \to \infty$, and a sequence of special functions $f_1(t), f_2(t), \cdots$ of the following description:

$$\begin{aligned} f_n(t) &= f(t) && \text{if } 0 \leq t \leq a_n, \\ 0 \leq f_n(t) &\leq f(t) && \text{if } a_n \leq t \leq a_{n+1}, \\ f_n(t) &= 0 && \text{if } a_{n+1} \leq t < \infty. \end{aligned} \tag{25}$$

Obviously, $f_n(t)$ increases monotonically to $f(t)$, and there exist a rational $\beta > 0$ and a sequence $\epsilon_n \searrow 0$ such that $f(t) - f_n(t) \leq \epsilon_n e^{-\beta t}$. Since

$$\int_0^{a_n} f(t)dE(t) \leq \int_0^\infty f_n(t)dE(t) = A(f_n) \leq \int_0^{a_{n+1}} f(t)dE(t) \leq A(f)$$

and $A(f) \leq A(f_n) + \epsilon_n T(\beta)$, we conclude, again on the basis of Lemma 1, that

$$\sup_a \int_0^a e^{-\alpha t}dE(t) \left(\equiv \int_0^\infty e^{-\alpha t}dE(t) \right) \tag{26}$$

exists and has the value $A(e^{-\alpha t}) = T(\alpha)$ for $\alpha > 0$, rational and irrational. This completes the proof of relation (2).

In the second place, if $E(t)$ is an increasing function from $0 \leq t < \infty$ to S, $E(0) = 0$, and if the quantity (26) exists for every $\alpha > 0$, then on the basis of the approximation (25) the quantity (24) exists for all $f(t) \geq 0$ which belong to F. Also, if $f_1 \geq 0, f_2 \geq 0, a_1 \geq 0, a_2 \geq 0, f(t) = a_1 f_1(t) + a_2 f_2(t)$, then $A(f) = a_1 A(f_1) + a_2 A(f_2)$. Now, if $f(t)$ is any element of F, then we may put $f = f^+ - f^-$, where $f^+(t) = \sup(f(t), 0)$, $f^-(t) = \sup(-f(t), 0)$, and for any other decomposition $f = f_1 - f_2, f_1 \geq 0, f_2 \geq 0$, we have $f_1 = f^+ + \varphi$, $f_2 = f^- + \varphi, \varphi \geq 0$.

This enables us to extend the definition of the integral (24), and hence of the distributive operation $A(f)$, to all of F, and the completed operation $A(f)$ is distributive and positive.

Finally, if a function $T(\alpha)$ is given as an integral (2), we may define two positive distributive operations. The first is a continuation from P to F of the operation $A'(p)$ as given by $A'(e^{-\alpha t}) = T(\alpha)$, $\alpha > 0$ rational. The second operation is given directly in terms of the integral (24). By the argument just completed, and by the uniqueness clause in Lemma 3, the two operations are identical. Hence we may assert that the quantity

$$A(\omega_{\rho\epsilon}(t)) = \int_0^\rho \omega_{\rho\epsilon}(t) dE(t)$$

is uniquely determined by the values of the function $T(\alpha)$. By our definition of integral

$$E(\rho - \epsilon) \leq A(\omega_{\rho\epsilon}(t)) \leq E(\rho).$$

Replacing ρ by $\rho - \epsilon$ $(0 < 2\epsilon < \rho)$, we have

$$E(\rho - 2\epsilon) \leq A(\omega_{\rho-\epsilon,\epsilon}(t)) \leq E(\rho - \epsilon),$$

and hence the quantity

$$E(\rho - 0) = \sup_\epsilon A(\omega_{\rho-\epsilon,\epsilon}(t)) \qquad (0 < \rho < \infty)$$

is also determined by $T(\alpha)$. Similarly

$$E(\rho + 0) = \inf_\epsilon A(\omega_{\rho+\epsilon,\epsilon}) \qquad (0 < \rho < \infty)$$

and $E(+0) = \inf_\epsilon E(\epsilon)$ are all determined in terms of $T(\alpha)$, and this completes the proof of the theorem.

Bibliography

1. F. Hausdorff, *Summationsmethoden und Momentfolgen.* I, Mathematische Zeitschrift, vol. 9(1921), pp. 98–99.
2. Einar Hille, *On semi-groups of transformations in Hilbert space*, Proceedings of the National Academy of Sciences, vol. 24(1938), pp. 159–161.
3. F. Riesz, *Ueber die linearen Transformationen des komplexen Hilbertschen Raumes*, Acta Szeged, vol. 5(1930), pp. 23–54.
4. D. V. Widder, *The Laplace Transform*, Princeton University Press, 1941, pp. 160–162.

Princeton University.

CONVERGENCE IN LENGTH AND CONVERGENCE IN AREA

By T. Radó and P. Reichelderfer

CHAPTER I

Introduction. Statement of Results.

1. We are given a sequence of functions $f_n(x)$ $(n = 0, 1, 2, \cdots)$, each defined on a closed (linear) interval $[a, b]$, and the following conditions.

(1.1) Each of the functions $f_n(x)$ $(n = 0, 1, 2, \cdots)$ is continuous on $[a, b]$.
(1.2) The functions $f_n(x)$ converge uniformly on $[a, b]$ to the function $f_0(x)$.
(1.3) Each of the functions $f_n(x)$ $(n = 0, 1, 2, \cdots)$ is of bounded variation on $[a, b]$. (Observe that this condition implies that the quantities involved in the remaining conditions in this section are finite.)
(1.4) The total variation $T(f_n)$ of $f_n(x)$ on $[a, b]$ converges to the total variation $T(f_0)$ of $f_0(x)$ on $[a, b]$.
(1.5) The total variation $T(f_n - f_0)$ of $f_n(x) - f_0(x)$ on $[a, b]$ converges to zero.
(1.6) The arc length $L(f_n)$ of the curve $y = f_n(x)$ $(a \leq x \leq b)$ converges to the arc length $L(f_0)$ of the curve $y = f_0(x)$ $(a \leq x \leq b)$.

Make the following definitions.

(1.7) A sequence of functions $f_n(x)$ satisfying conditions (1.1)-(1.4) is said to converge in variation to the function $f_0(x)$ on $[a, b]$—briefly, $f_n - v \to f_0$ on $[a, b]$.
(1.8) A sequence of functions $f_n(x)$ satisfying conditions (1.1)-(1.3), (1.5) is said to converge strongly in variation to the function $f_0(x)$ on $[a, b]$—briefly, $f_n - sv \to f_0$ on $[a, b]$.
(1.9) A sequence of functions $f_n(x)$ satisfying conditions (1.1)-(1.3), (1.6) is said to converge in length to the function $f_0(x)$ on $[a, b]$—briefly, $f_n - l \to f_0$ on $[a, b]$.
Since $| T(f_n) - T(f_0) | \leq T(f_n - f_0)$, it follows that $f_n - sv \to f_0$ on $[a, b]$ implies that $f_n - v \to f_0$ on $[a, b]$.

2. The terminology set forth in the preceding section is due to Adams, Clarkson, and Lewy [1], [2], but the definitions are not phrased as generally as they have given them. They do not require that the functions be continuous, or that the convergence be uniform. We have added these conditions because we want to think of these functions as representing continuous curves. Many of the theorems which we shall derive, together with the proofs which we give for them, are clearly valid if the continuity condition I, (1.1), that is, Chapter I, statement (1.1), is discarded, and if uniform convergence in condition I, (1.2) is replaced by ordinary convergence; in particular, this is true of the theorems which include,

Received March 4, 1942. Some of the results in this paper were presented to the American Mathematical Society at its meeting in Washington in May, 1941, by the senior author.

as special cases, the results of Adams and Lewy which we cite in the next section. However, continuity and uniform convergence will be essential conditions for other results in this paper.

3. Adams and Lewy [2] have established the following results. (We do not quote Adams and Lewy exactly, since we have modified their definitions (see I, §2).)

(3.1) If $f_n - l \to f_0$ on $[a, b]$, then $f_n - v \to f_0$ on $[a, b]$, but the converse is generally false even if the functions $f_n(x)$ $(n = 0, 1, 2, \cdots)$ are absolutely continuous on $[a, b]$.
(3.2) If $f_n - l \to f_0$ on $[a, b]$, then it is not generally true that $f_n - sv \to f_0$ on $[a, b]$. But if $f_n - l \to f_0$ on $[a, b]$, and if $f_0(x)$ is absolutely continuous on $[a, b]$, then $f_n - sv \to f_0$ on $[a, b]$.

4. The purpose of this paper is the investigation of generalizations which arise by replacing, in the definitions given in I, §1 and in the results cited in I, §3, non-parametric representations of plane curves by both parametric and non-parametric representations for space curves and for curved surfaces. In fact, we were interested in using, not the representations for curves or surfaces, but the curves or surfaces themselves in formulating our definitions and theorems. The nature of this question necessitated rather complicated geometrical considerations in order to insure that the concepts of length, area, and variation which we used belonged to the curve or surface and not to a particular representation thereof. This investigation we have carried out, but in order to reveal clearly the interesting analysis to which this question has led us, we choose to suppress bulky geometrical considerations and to work always, not with a curve or a surface, but with a particular representation for a curve or surface. On another occasion, we plan to treat the geometrical side of this question. Our results flow from general theorems in analysis. For this reason, it will be economical merely to state our definitions and results in this chapter. In Chapter II we shall present systematically general theorems in analysis, which we apply in Chapter III to make the proofs for our results.

5. Let $x(u)$, $y(u)$, $z(u)$ be a triple of real-valued functions defined and continuous on the closed (linear) interval $[\alpha, \beta]$. Then the equations

$$x = x(u), \qquad y = y(u), \qquad z = z(u) \qquad (\alpha \leq u \leq \beta)$$

give a parametric representation for a continuous curve in xyz-space. (For precise definitions of continuous curves and continuous surfaces, see [8], [10].) For conciseness, we employ triple notation and write these equations in the form

$$(5.1) \qquad \mathfrak{x} = \mathfrak{x}(u), \qquad u \in [\alpha, \beta] \qquad (\mathfrak{x}(u) = (x(u), y(u), z(u)); \alpha \leq u \leq \beta).$$

Corresponding representations for the projection of this curve upon the coordinate planes, yz, zx, xy, respectively, are given by the triples

$$(5.2)\quad x\text{-}\mathfrak{x}(u) = (0, y(u), z(u)), \qquad y\text{-}\mathfrak{x}(u) = (x(u), 0, z(u)),$$
$$z\text{-}\mathfrak{x}(u) = (x(u), y(u), 0) \quad (\alpha \leq u \leq \beta).$$

That is, the triples obtained by replacing the first, second and third components of $\mathfrak{x}(u)$ by zero will be denoted by $x\text{-}\mathfrak{x}(u)$, $y\text{-}\mathfrak{x}(u)$, $z\text{-}\mathfrak{x}(u)$, respectively.

6. For the convenience of the reader, we summarize some known results concerning continuous curves, which we have occasion to use in the sequel [10]. We assume that the reader is familiar with the notion of the arc length $L(\mathfrak{x})$ of the curve given by I, (5.1). A necessary and sufficient condition that $L(\mathfrak{x})$ be finite is that each of the functions $x(u)$, $y(u)$, $z(u)$ be of bounded variation on $[\alpha, \beta]$. If $L(\mathfrak{x})$ is finite, then the derivatives $x'(u)$, $y'(u)$, $z'(u)$ of $x(u)$, $y(u)$, $z(u)$, respectively, exist almost everywhere in $[\alpha, \beta]$, are summable, and

$$L(\mathfrak{x}) \geq \int_\alpha^\beta [x'(u)^2 + y'(u)^2 + z'(u)^2]^{\frac{1}{2}}\, du.$$

A necessary and sufficient condition that the sign of equality hold here is that each of the functions $x(u)$, $y(u)$, $z(u)$ be absolutely continuous on $[\alpha, \beta]$. If $\mathfrak{x}_n(u) = (x_n(u), y_n(u), z_n(u))$ is a sequence of triples of continuous functions such that $x_n(u)$, $y_n(u)$, $z_n(u)$ converge uniformly on $[\alpha, \beta]$ to $x(u)$, $y(u)$, $z(u)$, respectively, then $\liminf L(\mathfrak{x}_n) \geq L(\mathfrak{x})$.

7. Now we give definitions for convergence in variation and in length for parametric representations of space curves, which are analogous to the definitions in I, §1 for non-parametric representations of plane curves. We are given a sequence of triples $\mathfrak{x}_n(u) = (x_n(u), y_n(u), z_n(u))$ $(n = 0, 1, 2, \cdots)$ of functions defined on a closed interval $[\alpha, \beta]$, and the following conditions.

(7.1) Each of the functions $x_n(u)$, $y_n(u)$, $z_n(u)$ $(n = 0, 1, 2, \cdots)$ is continuous on $[\alpha, \beta]$.
(7.2) The functions $x_n(u)$, $y_n(u)$, $z_n(u)$ converge uniformly on $[\alpha, \beta]$ to $x_0(u)$, $y_0(u)$, $z_0(u)$, respectively.
(7.3) Each of the functions $x_n(u)$, $y_n(u)$, $z_n(u)$ $(n=0, 1, 2, \cdots)$ is of bounded variation on $[\alpha, \beta]$.
(7.4) The total variations $T(x_n)$, $T(y_n)$, $T(z_n)$ of $x_n(u)$, $y_n(u)$, $z_n(u)$, respectively, on $[\alpha, \beta]$ converge to the total variations $T(x_0)$, $T(y_0)$, $T(z_0)$ of $x_0(u)$, $y_0(u)$, $z_0(u)$, respectively, on $[\alpha, \beta]$.
(7.5) The total variations $T(x_n - x_0)$, $T(y_n - y_0)$, $T(z_n - z_0)$ of $x_n(u) - x_0(u)$, $y_n(u) - y_0(u)$, $z_n(u) - z_0(u)$, respectively, on $[\alpha, \beta]$ converge to zero.
(7.6) The arc length $L(\mathfrak{x}_n)$ of the curve represented by $\mathfrak{x} = \mathfrak{x}_n(u)$, $u \in [\alpha, \beta]$, converges to the arc length $L(\mathfrak{x}_0)$ of the curve represented by $\mathfrak{x} = \mathfrak{x}_0(u)$, $u \in [\alpha, \beta]$.

Make the following definitions.

(7.7) A sequence of triples $\mathfrak{x}_n(u)$ satisfying conditions (7.1)-(7.4) is said to *converge in variation* to the triple $\mathfrak{x}_0(u)$ on $[\alpha, \beta]$—briefly, $\mathfrak{x}_n - v \to \mathfrak{x}_0$ on $[\alpha, \beta]$.
(7.8) A sequence of triples $\mathfrak{x}_n(u)$ satisfying conditions (7.1)-(7.3), (7.5) is said to *converge strongly in variation* to the triple $\mathfrak{x}_0(u)$ on $[\alpha, \beta]$—briefly, $\mathfrak{x}_n - sv \to \mathfrak{x}_0$ on $[\alpha, \beta]$.
(7.9) A sequence of triples $\mathfrak{x}_n(u)$ satisfying conditions (7.1)-(7.3), (7.6) is said to *converge in length* to the triple $\mathfrak{x}_0(u)$ on $[\alpha, \beta]$—briefly, $\mathfrak{x}_n - l \to \mathfrak{x}_0$ on $[\alpha, \beta]$.

It is clear (cf. I, §1) that $\mathfrak{x}_n - sv \to \mathfrak{x}_0$ on $[\alpha, \beta]$ implies that $\mathfrak{x}_n - v \to \mathfrak{x}_0$ on $[\alpha, \beta]$.

8. Generalizing the results of Adams and Lewy cited in I, §3, we prove (see III, §§1, 2) the

THEOREM.

(8.1) *If* $\mathfrak{x}_n - l \to \mathfrak{x}_0$ *on* $[\alpha, \beta]$, *then* (*see* I, (5.2)) $x\text{-}\mathfrak{x}_n - l \to x\text{-}\mathfrak{x}_0$, $y\text{-}\mathfrak{x}_n - l \to y\text{-}\mathfrak{x}_0$, $z\text{-}\mathfrak{x}_n - l \to z\text{-}\mathfrak{x}_0$ *on* $[\alpha, \beta]$.
(8.2) *If* $\mathfrak{x}_n - l \to \mathfrak{x}_0$ *on* $[\alpha, \beta]$, *then* $\mathfrak{x}_n - v \to \mathfrak{x}_0$ *on* $[\alpha, \beta]$.
(8.3) *If* $\mathfrak{x}_n - l \to \mathfrak{x}_0$ *on* $[\alpha, \beta]$, *it does not generally follow that* $\mathfrak{x}_n - sv \to \mathfrak{x}_0$ *on* $[\alpha, \beta]$, *even though the functions* $x_n(u)$, $y_n(u)$, $z_n(u)$ $(n = 0, 1, 2, \cdots)$ *are absolutely continuous on* $[\alpha, \beta]$.
(8.4) *If* $\mathfrak{x}_n - sv \to \mathfrak{x}_0$ *on* $[\alpha, \beta]$, *and if* $x_n(u)$, $y_n(u)$, $z_n(u)$ *are absolutely continuous on* $[\alpha, \beta]$ *for* $n = 1, 2, \cdots$, *then* $x_0(u)$, $y_0(u)$, $z_0(u)$ *are absolutely continuous on* $[\alpha, \beta]$.

(It follows readily from the results of Adams and Lewy cited in I, (3.1) that no converse may be expected for either I, (8.1) or I, (8.2). For I, (8.4), we actually prove a sharper result (see III, §2).)

If we observe that the total variation of a single function, say $x(u)$ on $[\alpha, \beta]$, may be interpreted as the length of the one-dimensional curve given by $x = x(u)$, $y = 0$, $z = 0$ $(\alpha \leq u \leq \beta)$, then the first two parts of this theorem may be expressed geometrically as follows. If a sequence of parametric representations for continuous curves C_n in xyz-space converges in length to the parametric representation for a continuous curve C_0, then the corresponding sequence of representations for the projection of C_n on any one of the coordinate planes converges in length to the corresponding representation for the projection of C_0 on that coordinate plane, and the corresponding sequence of representations for the projection of C_n on any one of the coordinate axes converges in length to the corresponding representation for the projection of C_0 on that axis. (In fact, this statement remains valid for projections on any plane in space, and on any line in space.)

9. Let $y(x)$, $z(x)$ be a pair of functions defined and continuous on the closed (linear) interval $[a, b]$. Then the equations

$$y = y(x), \qquad z = z(x) \qquad (a \leq x \leq b)$$

give a non-parametric representation for a continuous curve in xyz-space. For conciseness, we write these equations in the form

$$\mathfrak{y} = \mathfrak{y}(x), \qquad x \in [a, b] \qquad (\mathfrak{y}(x) = (y(x), z(x)); a \leq x \leq b). \tag{9.1}$$

We denote the length of this curve by $L(\mathfrak{y})$. Remarks similar to those in I, §6 are valid. The reader will have no difficulty in formulating, for non-parametric representations of curves, definitions of convergence in variation, strong convergence in variation, and convergence in length, analogous to those given in I, §7 for parametric representations of curves.

10. Generalizing the results of Adams and Lewy cited in I, §3, we prove (see III, §3) the

THEOREM.

(10.1) *If* $\mathfrak{y}_n - l \to \mathfrak{y}_0$ *on* $[a, b]$, *then* (*see* I, §1) $y_n - l \to y_0$, $z_n - l \to z_0$ *on* $[a, b]$.
(10.2) *If* $\mathfrak{y}_n - l \to \mathfrak{y}_0$ *on* $[a, b]$, *then* $\mathfrak{y}_n - v \to \mathfrak{y}_0$ *on* $[a, b]$.
(10.3) *If* $\mathfrak{y}_n - l \to \mathfrak{y}_0$ *on* $[a, b]$ *and if* $y_0(x)$ *and* $z_0(x)$ *are both absolutely continuous on* $[a, b]$, *then* $\mathfrak{y}_n - sv \to \mathfrak{y}_0$ *on* $[a, b]$.
(10.4) *If* $\mathfrak{y}_n - l \to \mathfrak{y}_0$ *on* $[a, b]$ *and if* $y_n(x)$ *and* $z_n(x)$ *are both absolutely continuous on* $[a, b]$ *for* $n = 1, 2, \cdots$, *then* $y_0(x)$ *and* $z_0(x)$ *are both absolutely continuous on* $[a, b]$ (*see remark at the end of* §8).

This theorem admits a geometrical interpretation analogous to that given in I, §8 for parametric representations of curves. (10.3) and (10.4) contain the fact that if $\mathfrak{y}_n - l \to \mathfrak{y}_0$ on $[a, b]$ and if an infinite subsequence of the pairs of functions $y_n(x)$ and $z_n(x)$ are both absolutely continuous on $[a, b]$, then a necessary and sufficient condition that $\mathfrak{y}_n - sv \to \mathfrak{y}_0$ on $[a, b]$ is that both $y_0(x)$ and $z_0(x)$ be absolutely continuous on $[a, b]$.

11. This concludes a statement of our results for continuous curves. As we turn to consider continuous surfaces, a new problem confronts us. We want to work with concepts for the area of a surface and for bounded variation and absolute continuity of a representation for the surface. Now the corresponding concepts for curves are quite generally accepted as well known among mathematicians. But for continuous surfaces, there is no general agreement on the meanings for these concepts [8], [10], [12]. The theory of continuous surfaces, together with the companion theory of multiple integrals in the calculus of variations, is far from being complete. However, for the special case of surfaces given in non-parametric representation, the theory is as complete as the theory of continuous curves. We shall discuss this special case first, deriving results as

complete as those we have for curves. In I, §§12-17, we summarize the important features of this theory for the convenience of the reader (see [13; V] for a presentation of this theory and for references to the literature).

12. Let $f(x, y)$ be a function defined and continuous on the (planar) interval $[a, b; c, d]$. (By the interval $[a, b; c, d]$ in the xy-plane, we mean the set of points (x, y) satisfying $a \leq x \leq b, c \leq y \leq d$.) For fixed y in $[c, d]$, denote by $V_x(y, f)$ the total variation of $f(x, y)$ as a function of x on the interval $[a, b]$. Since $f(x, y)$ is continuous, it follows that $V_x(y, f)$ is a lower semi-continuous function on $[c, d]$. Define $V_y(x, f)$ by interchanging the rôles of x and y. If both $V_x(y, f)$ and $V_y(x, f)$ are summable on their respective intervals of definition, then $f(x, y)$ is said to be of bounded variation in the sense of Tonelli on the interval $[a, b; c, d]$ —briefly, BVT on $[a, b; c, d]$; we define the x- and y-variations of $f(x, y)$ on $[a, b; c, d]$ to be, respectively,

$$T_x(f) = \int_c^d V_x(y, f)\, dy, \qquad T_y(f) = \int_a^b V_y(x, f)\, dx.$$

13. Let $f(x, y)$ be a function defined and continuous on the interval $[a, b; c, d]$. Denote by $E(y, f)$ the set of points y contained in $[c, d]$ for each of which $f(x, y)$ is an absolutely continuous function of x on the interval $[a, b]$. Since $f(x, y)$ is continuous, it follows that $E(y, f)$ is a Borel set, hence measurable. Define $E(x, f)$ by interchanging the rôles of x and y. If $f(x, y)$ is BVT on $[a, b; c, d]$, and if

$$| E(y, f) | = d - c, \qquad | E(x, f) | = b - a,$$

then $f(x, y)$ is said to be absolutely continuous in the sense of Tonelli on $[a, b; c, d]$—briefly, ACT on $[a, b; c, d]$. (If E be any point set in n-dimensional Euclidean space, then $| E |$ denotes the n-dimensional exterior measure of E.)

14. Let $f(x, y)$ be a function defined and continuous on the interval $[a, b; c, d]$. For fixed y in $[c, d]$, denote by $L_x(y, f)$ the length of the curve $z = f(x, y)$ $(a \leq x \leq b)$. Since $f(x, y)$ is continuous, it follows that $L_x(y, f)$ is a lower semi-continuous function on $[c, d]$; thus $L_x(y, f)$ is a non-negative, measurable function. Define $L_y(x, f)$ by interchanging the rôles of x and y. Clearly

$$V_x(y, f) \leq L_x(y, f) \leq b - a + V_x(y, f), \qquad y \in [c, d],$$

$$V_y(x, f) \leq L_y(x, f) \leq d - c + V_y(x, f), \qquad x \in [a, b],$$

so that a necessary and sufficient condition for the summability of $L_x(y, f)$ and $L_y(x, f)$ is that $f(x, y)$ be BVT on $[a, b; c, d]$. If $f(x, y)$ is BVT on $[a, b; c, d]$, define

$$S_x(f) = \int_c^d L_x(y, f)\, dy, \qquad S_y(f) = \int_a^b L_y(x, f)\, dx.$$

15. Let $f(x, y)$ be a function defined and continuous on the interval $[a, b; c, d]$. Then the equation

$$z = f(x, y) \qquad (a \leq x \leq b, c \leq y \leq d) \tag{15.1}$$

gives a non-parametric representation for a continuous surface [5], [6]. If it is possible to subdivide the interval $[a, b; c, d]$ into a finite number of non-overlapping triangles, on each of which $f(x, y)$ is a linear function in both x and y, then $f(x, y)$ is said to be quasi-linear on $[a, b; c, d]$, and the surface represented by (15.1) is termed a polyhedron. Now the image of a triangle on which $f(x, y)$ is linear in both x and y is again a (possibly degenerate) triangle; thus a polyhedron is composed of a finite number of triangles which are the respective images of the triangles in a subdivision of the interval $[a, b; c, d]$ into non-overlapping triangles. The sum of the areas of these image triangles is the elementary area $a(f)$ of the polyhedron, and is independent of the choice of subdivision of $[a, b; c, d]$.

16. The Lebesgue area $A(f)$ of the continuous surface given by I, (15.1) may be defined as follows. (This is not the original definition for the Lebesgue area of a continuous surface [8], [10]. However, this definition is equivalent to the Lebesgue definition, for surfaces in non-parametric representation.) Let $f_n(x, y)$ $(n = 1, 2, \cdots)$ be a sequence of quasi-linear functions on the interval $[a, b; c, d]$ such that $f_n(x, y)$ converges uniformly on $[a, b; c, d]$ to $f(x, y)$. Then $\liminf a(f_n)$ is an upper bound for the Lebesgue area $A(f)$; and $A(f)$ is the greatest lower bound of all the upper bounds obtained in this way. If $f(x, y)$ is quasi-linear on $[a, b; c, d]$, one has $A(f) = a(f)$.

17. Let I, (15.1) be a non-parametric representation for a continuous surface. A necessary and sufficient condition that the area $A(f)$ be finite is that $f(x, y)$ be BVT on $[a, b; c, d]$. If $A(f)$ is finite, then the partial derivatives

$$p(x, y) = \frac{\partial f}{\partial x}, \qquad q(x, y) = \frac{\partial f}{\partial y}$$

exist almost everywhere in the interval $[a, b; c, d]$ and are summable; and (see I, §§12-16)

$$T_x(f) \geq \int_a^b \int_c^d |\, p(x, y)\,|\, dx\, dy, \qquad T_y(f) \geq \int_a^b \int_c^d |\, q(x, y)\,|\, dx\, dy; \tag{17.1}$$

$$S_x(f) \geq \int_a^b \int_c^d [p(x, y)^2 + 1]^{\frac{1}{2}}\, dx\, dy,$$

$$S_y(f) \geq \int_a^b \int_c^d [q(x, y)^2 + 1]^{\frac{1}{2}}\, dx\, dy; \tag{17.2}$$

$$A(f) \geq \int_a^b \int_c^d [p(x, y)^2 + q(x, y)^2 + 1]^{\frac{1}{2}}\, dx\, dy. \tag{17.3}$$

A necessary and sufficient condition that the signs of equality hold in relations (17.1), (17.2) or (17.3) is that $f(x, y)$ be ACT on $[a, b; c, d]$. If $f_n(x, y)$ be a sequence of continuous functions such that $f_n(x, y)$ converges uniformly on $[a, b; c, d]$ to $f(x, y)$, then

$$\text{(17.4)} \qquad \liminf T_x(f_n) \geq T_x(f), \qquad \liminf T_y(f_n) \geq T_y(f);$$

$$\text{(17.5)} \qquad \liminf S_x(f_n) \geq S_x(f), \qquad \liminf S_y(f_n) \geq S_y(f);$$

$$\text{(17.6)} \qquad \liminf A(f_n) \geq A(f).$$

If $f_1(x, y)$, $f_2(x, y)$ be defined and continuous on $[a, b; c, d]$, then

$$\text{(17.7)} \quad T_x(f_1 \pm f_2) \leq T_x(f_1) + T_x(f_2), \qquad T_y(f_1 \pm f_2) \leq T_y(f_1) + T_y(f_2).$$

18. Now we give definitions for convergence in variation and in area for non-parametric representations of surfaces, which are analogous to those in I, §1 for non-parametric representations of plane curves. We are given a sequence of functions $f_n(x, y)$ $(n = 0, 1, 2, \cdots)$ each defined on a closed interval $[a, b; c, d]$, and the following conditions.

(18.1) Each of the functions $f_n(x, y)$ $(n = 0, 1, 2, \cdots)$ is continuous on $[a, b; c, d]$.
(18.2) The functions $f_n(x, y)$ converge uniformly on $[a, b; c, d]$ to the function $f_0(x, y)$.
(18.3) Each of the functions $f_n(x, y)$ $(n = 0, 1, 2, \cdots)$ is BVT on $[a, b; c, d]$.
(18.4) The x- and y-variations, $T_x(f_n)$ and $T_y(f_n)$, converge to $T_x(f_0)$ and $T_y(f_0)$ respectively.
(18.5) The x- and y-variations, $T_x(f_n - f_0)$ and $T_y(f_n - f_0)$, converge to zero.
(18.6) The areas $A(f_n)$ of the surfaces represented by $z = f_n(x, y)$, $(x, y) \in [a, b; c, d]$, converge to the area $A(f_0)$ of the surface represented by $z = f_0(x, y)$, $(x, y) \in [a, b; c, d]$.

Make the following definitions.

(18.7) A sequence of functions $f_n(x, y)$ satisfying conditions (18.1)-(18.4) is said to *converge in variation* to the function $f_0(x, y)$ on $[a, b; c, d]$—briefly, $f_n - v \to f_0$ on $[a, b; c, d]$.
(18.8) A sequence of functions $f_n(x, y)$ satisfying conditions (18.1)-(18.3), (18.5) is said to *converge strongly in variation* to the function $f_0(x, y)$ on $[a, b; c, d]$—briefly, $f_n - sv \to f_0$ on $[a, b; c, d]$.
(18.9) A sequence of functions $f_n(x, y)$ satisfying conditions (18.1)-(18.3), (18.6) is said to *converge in area* to the function $f_0(x, y)$ on $[a, b; c, d]$—briefly, $f_n - a \to f_0$ on $[a, b; c, d]$.

Since, from I, (17.7), it is clear that

$$| T_x(f_n) - T_x(f_0) | \leq T_x(f_n - f_0), \qquad | T_y(f_n) - T_y(f_0) | \leq T_y(f_n - f_0),$$

it follows that $f_n - sv \to f_0$ on $[a, b; c, d]$ implies that $f_n - v \to f_0$ on $[a, b; c, d]$. From I, (17.4)-(17.6), it follows that if f_n converges in variation, strongly in variation, or in area, to f_0 on $[a, b; c, d]$ it does likewise on any subinterval of $[a, b; c, d]$. Moreover, if $f_n(x, y)$ $(n = 0, 1, 2, \cdots)$ are BVT on $[a, b; c, d]$, then (see I, (17.1))

$$T_x(f_n - f_0) \geq \int_a^b \int_c^d | p_n(x, y) - p_0(x, y) | \, dx \, dy,$$

$$T_y(f_n - f_0) \geq \int_a^b \int_c^d | q_n(x, y) - q_0(x, y) | \, dx \, dy.$$

From elementary inequalities, we derive the

COROLLARY. *If* $f_n - sv \to f_0$ *on* $[a, b; c, d]$, *then*

$$\lim \int_a^b \int_c^d [\{ p_n(x, y) - p_0(x, y) \}^2 + \{ q_n(x, y) - q_0(x, y) \}^2]^{\frac{1}{2}} \, dx \, dy = 0.$$

19. Generalizing the results of Adams and Lewy cited in I, §3, we prove (see III, §§4-6) the

THEOREM.

(19.1) *If* $f_n - a \to f_0$ *on* $[a, b; c, d]$, *then* (*see* I, §14) $S_x(f_n)$ *and* $S_y(f_n)$ *converge to* $S_x(f_0)$ *and* $S_y(f_0)$ *respectively.*
(19.2) *If* $f_n - a \to f_0$ *on* $[a, b; c, d]$, *then* $f_n - v \to f_0$ *on* $[a, b; c, d]$.
(19.3) *If* $f_n - a \to f_0$ *on* $[a, b; c, d]$, *then it does not generally follow that* $f_n - sv \to f_0$ *on* $[a, b; c, d]$, *but one has* (*see* I, §§12, 13)

$$\lim \int_{E(y, f_0)} V_x(y, f_n - f_0) \, dy = 0, \qquad \lim \int_{E(x, f_0)} V_y(x, f_n - f_0) \, dx = 0.$$

Thus (*see* I, §§12, 13), *if* $f_n - a \to f_0$ *on* $[a, b; c, d]$ *and if* $f_0(x, y)$ *is* ACT *on* $[a, b; c, d]$, *then* $f_n - sv \to f_0$ *on* $[a, b; c, d]$.
(19.4) *If* $f_n - sv \to f_0$ *on* $[a, b; c, d]$ *and if* $f_n(x, y)$ *is* ACT *on* $[a, b; c, d]$ *for* $n = 1, 2, \cdots$, *then* $f_0(x, y)$ *is also* ACT *on* $[a, b; c, d]$ (*see* I, §20).

We give a direct proof of the last statement in I, (19.3) in III, §6. A stronger statement than I, (19.4) is actually true (see I, §20; cf. I, §8).

20. In order that the reader may better understand some of the implications of this theorem, we pause to review some of the results in the literature. McShane [4; Theorem V] showed that if $f_n - a \to f_0$ on $[a, b; c, d]$, then $[(p_n - p_0)^2 + (q_n - q_0)^2]^{\frac{1}{2}}$ converges to zero in measure on $[a, b; c, d]$ (see II, §16). Radó and Reichelderfer [11] improved this result by showing that, if $f_n - a \to f_0$ on $[a, b; c, d]$, then

$$\lim \int_a^b \int_c^d [\{p_n(x, y) - p_0(x, y)\}^2 + \{q_n(x, y) - q_0(x, y)\}^2]^{\lambda/2} \, dx \, dy = 0 \tag{20.1}$$

for every exponent λ satisfying $0 < \lambda < 1$. Now McShane [4; Theorem VI] has shown that if a sequence of functions $f_n(x, y)$ $(n = 0, 1, 2, \cdots)$ satisfies conditions I, (18.1)-(18.3) and if, moreover, each $f_n(x, y)$ is ACT on $[a, b; c, d]$ for $n = 1, 2, \cdots$, then

$$\liminf \int_a^b \int_c^d [\{p_n - p_0\}^2 + \{q_n - q_0\}^2]^{\frac{1}{2}} \, dx \, dy \geq L(f_0) - \int_a^b \int_c^d [p_0(x, y)^2 + q_0(x, y)^2 + 1]^{\frac{1}{2}} \, dx \, dy.$$

From the facts stated in I, §17, it follows that the sign of equality cannot hold in (20.1) for the exponent $\lambda = 1$ unless $f_0(x, y)$ is ACT on $[a, b; c, d]$. (19.4) follows at once from the corollary in I, §18 and from this fact. (19.3) and (19.4) thus imply the fact that if $f_n - a \to f_0$ on $[a, b; c, d]$ and if an infinite subsequence of the $f_n(x, y)$ are ACT on $[a, b; c, d]$, then a necessary and sufficient condition that $f_n - sv \to f_0$ on $[a, b; c, d]$ is that $f_0(x, y)$ be ACT on $[a, b; c, d]$ (cf. I, §10).

21. Turning now to the case of parametric representations of continuous surfaces, we confront the problems already mentioned in I, §11. Since there is no general agreement on the notions for the area of a surface given in parametric representation, or for bounded variation and absolute continuity of a representation for that surface, we may be justified in stating what we regard as a useful principle for the purpose of directing work in this field.

Consider a continuous surface given in parametric representation

$$x = x(u, v), \qquad y = y(u, v), \qquad z = z(u, v) \qquad (\alpha \leq u \leq \beta, \gamma \leq v \leq \delta).$$

For brevity, we use triple notation, and write these equations in the form

$$\mathfrak{x} = \mathfrak{x}(u, v), \qquad (u, v) \in \mathfrak{I} \tag{21.1}$$
$$(\mathfrak{x}(u, v) = (x(u, v), y(u, v), z(u, v)); (u, v) \in \mathfrak{I} = [\alpha, \beta; \gamma, \delta]).$$

Make the following assumptions.

(21.2) Some definition is given for the area of the surface represented by equations (21.1); denote this area by $\mathfrak{A}(\mathfrak{x})$ (for examples, see [8]).
(21.3) For each of the pairs of functions

$$(y(u, v), z(u, v)), \qquad (z(u, v), x(u, v)), \qquad (x(u, v), y(u, v)) \qquad ((u, v) \in \mathfrak{I})$$

some definition is given for Jacobians $\mathfrak{J}(x; u, v)$, $\mathfrak{J}(y; u, v)$, $\mathfrak{J}(z; u, v)$, respectively (for examples, see [8], [12]). Then the representation (21.1) is said to be *absolutely continuous* $\mathfrak{A}\mathfrak{J}$ provided

(21.4) each of the Jacobians $\mathfrak{J}(x; u, v)$, $\mathfrak{J}(y; u, v)$, $\mathfrak{J}(z; u, v)$ exists almost everywhere and is summable on $\mathfrak{I}$;

$$(21.5) \qquad \mathfrak{A}(\mathfrak{x}) = \iint_{\mathfrak{I}} [\mathfrak{J}(x; u, v)^2 + \mathfrak{J}(y; u, v)^2 + \mathfrak{J}(z; u, v)^2]^{\frac{1}{2}}\, du\, dv.$$

The reader will find, if he applies the analogous principle for continuous curves, using the usual definition of length and the ordinary derivatives of the representing functions, that he obtains a criterion for an absolutely continuous representation of a continuous curve which is equivalent to the usual definition (see I, §6). We now present a few results for parametric representations of continuous surfaces. First, we agree on an area and a Jacobian.

22. We shall use the Lebesgue area of the surface given by the equations I, (21.1), which may be described as follows. (This is not the original definition for the Lebesgue area of a continuous surface [8], [10], but it is an equivalent definition.) First, let us assume that each of the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ is quasi-linear on $\mathfrak{I}$ (see I, §15); then the surface represented by I, (21.1) is termed a polyhedron. Consider any subdivision of $\mathfrak{I}$ into non-overlapping triangles on each of which each of the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ is linear in both u and v. The image of any triangle in this subdivision is a (possibly degenerate) triangle. The sum of the areas of the images of these triangles is the elementary area $a(\mathfrak{x})$ of the polyhedron; it is independent of the particular subdivision of $\mathfrak{I}$ chosen.

23. Now let the equations I, (21.1) represent a general continuous surface. Assume that

$$\mathfrak{x}_n(u, v) = (x_n(u, v), y_n(u, v), z_n(u, v)) \qquad ((u, v) \in \mathfrak{I};\ n = 1, 2, 3, \cdots)$$

is a sequence of triples of quasi-linear functions for which $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$ converge uniformly on $\mathfrak{I}$ to $x(u, v)$, $y(u, v)$, $z(u, v)$, respectively. Then $\liminf a(\mathfrak{x}_n)$ is an upper bound for the Lebesgue area $A(\mathfrak{x})$ of the surface given

by I, (21.1), and $A(\mathfrak{x})$ is the greatest lower bound of all the upper bounds obtained in this way. If the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ are quasi-linear on $\mathfrak{I}$, then $A(\mathfrak{x}) = a(\mathfrak{x})$. Let

$$\mathfrak{x}_n(u, v) = (x_n(u, v), y_n(u, v), z_n(u, v)) \qquad ((u, v) \in \mathfrak{I};\ n = 1, 2, 3, \cdots)$$

be a sequence of triples of continuous functions such that

(23.1) the functions $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$ converge uniformly on $\mathfrak{I}$ to $x(u, v)$, $y(u, v)$, $z(u, v)$, respectively.

Then it is true that $\liminf A(\mathfrak{x}_n) \geq A(\mathfrak{x})$.

24. We shall use the ordinary Jacobians, defined by

$$J(x; u, v) = \partial(y, z)/\partial(u, v), \qquad J(y; u, v) = \partial(z, x)/\partial(u, v),$$
$$J(z; u, v) = \partial(x, y)/\partial(u, v),$$

wherever the derivatives involved exist. If these Jacobians exist almost everywhere on the interior of $\mathfrak{I}$ and if the area $A(\mathfrak{x})$ is finite, then Radó [7] has proved that these Jacobians are summable on $\mathfrak{I}$, and

$$A(\mathfrak{x}) \geq \iint_{\mathfrak{I}} [J(x; u, v)^2 + J(y; u, v)^2 + J(z; u, v)^2]^{\frac{1}{2}}\, du\, dv.$$

Furthermore, a necessary and sufficient condition that the sign of equality hold in this inequality—that is, that $\mathfrak{x}(u, v)$ be absolutely continuous AJ (see I, §21)—is that there exist a sequence of triples $\mathfrak{x}_n(u, v)$ of continuous functions satisfying condition I, (23.1) for which

$$\limsup A(\mathfrak{x}_n) \leq \iint_{\mathfrak{I}} [J(x; u, v)^2 + J(y; u, v)^2 + J(z; u, v)^2]^{\frac{1}{2}}\, du\, dv.$$

25. Here is our main result for parametric representations of surfaces (see III, §§14, 15).

THEOREM. *Consider a triple of continuous functions*

(25.1) $\qquad \mathfrak{x}(u, v) = (x(u, v), y(u, v), z(u, v)) \qquad ((u, v) \in \mathfrak{I} = [\alpha, \beta; \gamma, \delta])$

for which
(25.2) *each of the Jacobians $J(x; u, v)$, $J(y; u, v)$, $J(z; u, v)$ exists almost everywhere in the interior of $\mathfrak{I}$ and*
(25.3) *the area $A(\mathfrak{x})$ is finite.*

Let

$$(25.4)\quad \mathfrak{x}_n(u, v) = (x_n(u, v), y_n(u, v), z_n(u, v)) \qquad ((u, v) \in \mathfrak{I};\ n = 1, 2, 3, \cdots)$$

be any sequence of continuous triples for which

(25.5) *the functions* $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$ *converge uniformly on* $\mathfrak{I}$ *to* $x(u, v)$, $y(u, v)$, $z(u, v)$, *respectively;*
(25.6) *each of the triples* $\mathfrak{x}_n(u, v)$, $n = 1, 2, \cdots$, *is absolutely continuous* AJ;
(25.7) *the area* $A(\mathfrak{x}_n)$ *converges to*

$$\iint_{\mathfrak{I}} [J(x;\ u,\ v)^2 + J(y;\ u,\ v)^2 + J(z;\ u,\ v)^2\ ]^{\frac{1}{2}}\, du\, dv.$$

Then

$$\lim \iint_{\mathfrak{I}} |\ J(x_n;\ u,\ v)\ |\, du\, dv = \iint_{\mathfrak{I}} |\ J(x;\ u,\ v)\ |\, du\, dv,$$

$$(25.8)\qquad \lim \iint_{\mathfrak{I}} |\ J(y_n;\ u,\ v)\ |\, du\, dv = \iint_{\mathfrak{I}} |\ J(y;\ u,\ v)\ |\, du\, dv,$$

$$\lim \iint_{\mathfrak{I}} |\ J(z_n;\ u,\ v)\ |\, du\, dv = \iint_{\mathfrak{I}} |\ J(z;\ u,\ v)\ |\, du\, dv.$$

From this theorem follow two important corollaries (see I, §§26, 27 and III, §§16, 17).

26. Corollary. *If the triple* $\mathfrak{x}(u, v)$ *is absolutely continuous* AJ, *then each of the triples*

$$(26.1)\quad x\text{-}\mathfrak{x}(u, v) = (0, y(u, v), z(u, v)), \qquad y\text{-}\mathfrak{x}(u, v) = (x(u, v), 0, z(u, v)),$$

$$z\text{-}\mathfrak{x}(u, v) = (x(u, v), y(u, v), 0) \qquad ((u, v) \in \mathfrak{I})$$

is also absolutely continuous AJ. (*See the remark at the end of* §5 *for the notation.*)

Geometrically, this corollary states that if a parametric representation for a continuous surface is absolutely continuous AJ, then the corresponding representation for the projection of that surface on any one of the coordinate planes is absolutely continuous AJ. (In fact, by further considerations of a geometrical nature, one may show that the corresponding representation for the projection of this surface on any plane is absolutely continuous.) This result has an analogue for continuous curves, which the reader may formulate.

27. The pairs of equations

$$\begin{cases} y = y(u, v), \\ z = z(u, v), \end{cases} \quad \begin{cases} z = z(u, v), \\ x = x(u, v), \end{cases} \quad \begin{cases} x = x(u, v), \\ y = y(u, v), \end{cases} \qquad (u, v) \in \mathfrak{I}, \tag{27.1}$$

may be regarded as defining continuous transformations from the interval $\mathfrak{I}$ to the yz-, zx-, xy-planes respectively. The absolute continuity of the type guaranteed by the preceding corollary tells us nothing as to what sort of transformation formulas, if any, we may expect for these transformations. It is interesting, and important for applications, to observe that we have another

COROLLARY. *If the triple $\xi(u, v)$ is absolutely continuous AJ, then each of the three transformations given by* (27.1) *is absolutely continuous in another sense—it belongs to the class K_3.*

This class K_3 has been defined and studied by the authors [12; §1.37]. (For simplicity, we have modified the notation slightly; in our earlier paper, this class is denoted by $K_3(\mathfrak{I}^0)$, where $\mathfrak{I}^0$ is the set of interior points of $\mathfrak{I}$.) In order to define this class, we would be forced to review a great number of definitions and results which are extraneous to the present question. We choose, therefore, to describe this class K_3 by stating some of its important properties, referring the reader to our earlier paper cited above for details.

28. If $x(u, v)$, $y(u, v)$ are any pair of real-valued functions, each defined and continuous on an interval $\mathfrak{I} = [\alpha, \beta; \gamma, \delta]$, then the relations

$$x = x(u, v), \qquad y = y(u, v) \qquad ((u, v) \in \mathfrak{I}) \tag{28.1}$$

define a continuous transformation T from the interval $\mathfrak{I}$ to a bounded portion of the xy-plane. (We do not mean to insist that the transformations T in the class K_3 must be defined from the uv-plane to the xy-plane; rather, we consider K_3 as the class of all flat continuous transformations which transform the interval $\mathfrak{I}$ into a portion of some plane in a certain way [12].) A necessary condition that T be in the class K_3 is that the ordinary Jacobian $J(u, v) = \partial(x, y)/\partial(u, v)$ exist almost everywhere on the interior of $\mathfrak{I}$ and be summable. If the functions $x(u, v)$, $y(u, v)$ have continuous partial derivatives of the first order in the interior of $\mathfrak{I}$, together with a summable Jacobian, then T is in the class K_3. If the functions $x(u, v)$, $y(u, v)$ both satisfy a Lipschitz condition in u and v, then the transformation T is in the class K_3. Thus the class K_3 contains many of the transformations which are of importance in the applications. Furthermore, for this class K_3, we have the following

CLOSURE THEOREM. *Let*

$$T: x = x(u, v), y = y(u, v), \qquad (u, v) \in \mathfrak{I};$$

$$T_n : x = x_n(u, v), y = y_n(u, v) \qquad ((u, v) \in \mathfrak{I}, n = 1, 2, 3, \cdots)$$

be continuous transformations with the following properties: the ordinary Jacobian $J(u, v) = \partial(x, y)/\partial(u, v)$ *exists almost everywhere in the interior of* $\mathfrak{I}$ *and is summable; each of the transformations* T_n *belongs to the class* K_3 *for* $n = 1, 2, 3, \cdots$; *the functions* $x_n(u, v)$, $y_n(u, v)$ *converge uniformly on* $\mathfrak{I}$ *to* $x(u, v)$, $y(u, v)$, *respectively; for every interval* Δ *contained in* $\mathfrak{I}$, *it is true that*

$$\lim \iint_{\Delta} | J_n(u, v) | \, du \, dv = \iint_{\Delta} | J(u, v) | \, du \, dv,$$

where the $J_n(u, v)$ *are the Jacobians for the transformations* T_n $(n = 1, 2, 3, \cdots)$. *Then* T *is also in the class* K_3.

This theorem is not quite the one given in [12; §1.41], the last condition of that theorem having been replaced by a weaker one. However, this theorem follows at once by a trivial modification of the proof given by Radó and Reichelderfer.

The importance of this class K_3 lies in the fact that, for transformations in it, transformation formulas of great generality are valid. For example [12; §1.34], suppose that T is any transformation in the class K_3 which transforms the boundary of the interval $\mathfrak{I}$ into a set of measure zero in the xy-plane. Denote by $\mu(x, y)$ the topological index of the point (x, y) with respect to the image of the boundary of $\mathfrak{I}$, provided this point is not on the image of the boundary of $\mathfrak{I}$; otherwise, put $\mu(x, y) = 0$. Then if $H(x, y)$ is any measurable function in the xy-plane, we have

$$\iint_{\mathfrak{I}} H(x(u, v), y(u, v)) \, J(u, v) \, du \, dv = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} H(x, y) \, \mu(x, y) \, dx \, dy,$$

as soon as the integral on the left exists.

CHAPTER II

General Lemmas

The lemmas in this chapter are stated in a form convenient for our purposes. Both the statements of these lemmas and their proofs permit substantial generalization.

1. Let $\mathfrak{I}$ be a (closed) interval in Euclidean n-space—that is, a set of points in Euclidean n-space whose coordinates $u = (u^1, \cdots, u^n)$ satisfy relations of the form $\alpha^1 \leq u^1 \leq \beta^1, \cdots, \alpha^n \leq u^n \leq \beta^n$, where $\alpha^1, \beta^1, \cdots, \alpha^n, \beta^n$ are real numbers

satisfying $\alpha^1 < \beta^1, \cdots, \alpha^n < \beta^n$. Let δ, Δ denote generic intervals contained in $\mathfrak{I}$. By a subdivision $D(\Delta)$ of an interval Δ contained in $\mathfrak{I}$, we mean a finite system of intervals δ satisfying (see I, end of §13),

$$\Delta = \sum_{\delta \in D(\Delta)} \delta, \qquad |\Delta| = \sum_{\delta \in D(\Delta)} |\delta|.$$

The maximum of the diameters of the intervals $\delta \in D(\Delta)$ is denoted by $||D(\Delta)||$.

2. By an interval function $\phi(\Delta)$, we mean a law which assigns to every interval Δ contained in $\mathfrak{I}$ a unique, finite, real number. If $\phi(\Delta)$ is a non-negative number for every interval Δ contained in $\mathfrak{I}$, then $\phi(\Delta)$ is said to be a non-negative interval function. For a subdivision $D(\Delta)$, we define

$$\phi(D(\Delta)) = \sum_{\delta \in D(\Delta)} \phi(\delta).$$

(More generally, if $E(\Delta)$ is any finite system of non-overlapping intervals δ contained in Δ, we put

$$\phi(E(\Delta)) = \sum_{\delta \in E(\Delta)} \phi(\delta).)$$

An interval function $\phi(\Delta)$ is said to increase (decrease) by subdivision if, for every interval Δ contained in $\mathfrak{I}$ and for every subdivision $D(\Delta)$, it is true that $\phi(\Delta) \leq \phi(D(\Delta))$ $(\phi(\Delta) \geq \phi(D(\Delta)))$. If an interval function both increases and decreases by subdivision, it is said to be additive.

3. Let $\phi(\Delta)$ be any interval function. If Δ be any interval contained in $\mathfrak{I}$, define

$$U(\Delta, \phi) = \text{l.u.b. } \phi(D(\Delta))$$

for all subdivisions $D(\Delta)$. If $U(\Delta, \phi)$ is finite for every interval Δ contained in $\mathfrak{I}$, we say that $\phi(\Delta)$ has a U-function. Then $U(\Delta, \phi)$ is an interval function, which decreases by subdivision. However, if $\phi(\Delta)$ increases by subdivision, then $U(\Delta, \phi)$ is additive.

4. If $\phi(\Delta)$, $\psi(\Delta)$ be any pair of interval functions, we define a non-negative interval function $\omega(\Delta)$ by the relation

$$\omega(\Delta) = [\phi(\Delta)^2 + \psi(\Delta)^2]^{\frac{1}{2}} \qquad (\Delta \subset \mathfrak{I}).$$

If both $\phi(\Delta)$ and $\psi(\Delta)$ are non-negative and increase by subdivision, then clearly $\omega(\Delta)$ increases by subdivision. If both $\phi(\Delta)$ and $\psi(\Delta)$ are non-negative and have U-functions, then $\omega(\Delta)$ has a U-function.

5. LEMMA. *Let $\phi_n(\Delta)$, $\psi_n(\Delta)$ $(n = 0, 1, 2, \cdots)$ be a sequence of pairs of interval functions satisfying the following conditions.*

(5.1) *The interval functions $\phi_n(\Delta)$ and $\psi_n(\Delta)$ are non-negative for $n = 0, 1, 2, \cdots$.*

$$\text{(5.2)} \qquad \liminf \phi_n(\Delta) \geq \phi_0(\Delta), \qquad \liminf \psi_n(\Delta) \geq \psi_0(\Delta), \qquad (\Delta \subset \mathfrak{I}).$$

Then

$$\text{(5.3)} \qquad \liminf \omega_n(\Delta) \geq \omega_0(\Delta) \qquad (\Delta \subset \mathfrak{I});$$

$$\text{(5.4)} \qquad \liminf U(\Delta, \phi_n) \geq U(\Delta, \phi_0); \qquad \liminf U(\Delta, \psi_n) \geq U(\Delta, \psi_0) \quad (\Delta \subset \mathfrak{I});$$

$$\text{(5.5)} \qquad \liminf U(\Delta, \omega_n) \geq U(\Delta, \omega_0) \qquad (\Delta \subset \mathfrak{I}).$$

Proof. From II, (5.1), (5.2) and the definition of $\omega_n(\Delta)$, relation II, (5.3) follows immediately. If $D(\Delta)$ be any subdivision, then

$$\liminf U(\Delta, \phi_n) \geq \liminf \phi_n(D(\Delta)) \geq \phi_0(D(\Delta));$$

consequently, $\liminf U(\Delta, \phi_n) \geq U(\Delta, \phi_0)$, and the first inequality in II, (5.4) is verified. The remaining inequalities follow by similar reasonings.

6. LEMMA. *Let $\phi_n(\Delta)$, $\psi_n(\Delta)$ $(n = 0, 1, 2, \cdots)$ be a sequence of pairs of interval functions satisfying* II, (5.1), (5.2) *and the following conditions.*

(6.1) *The interval functions $\phi_n(\Delta)$ and $\psi_n(\Delta)$ increase by subdivision for $n = 0, 1, 2, \cdots$.*

(6.2) *The interval functions $\phi_n(\Delta)$ and $\psi_n(\Delta)$ have U-functions for $n = 0, 1, 2, \cdots$.*

$$\text{(6.3)} \qquad \lim U(\mathfrak{I}, \omega_n) = U(\mathfrak{I}, \omega_0).$$

Then

$$\text{(6.4)} \qquad \lim U(\Delta, \omega_n) = U(\Delta, \omega_0) \qquad (\Delta \subset \mathfrak{I}).$$

Proof. If Δ_0 be any interval in $\mathfrak{I}$, it is clear that there exists a subdivision $D(\mathfrak{I})$ for which Δ_0 is an element (see II, §1). From II, (5.1), (6.1), (6.2), it follows that $U(\Delta, \omega_n)$ is additive for $n = 0, 1, 2, \cdots$ (see II, §§3, 4). Thus, from II, (6.3), (5.5) follows

$$\begin{aligned} U(\mathfrak{I}, \omega_0) = \lim U(\mathfrak{I}, \omega_n) &= \limsup U(D(\mathfrak{I}), \omega_n) \\ &\geq \limsup U(\Delta_0, \omega_n) + \liminf U(D(\mathfrak{I}) - \Delta_0, \omega_n) \\ &\geq \liminf U(\Delta_0, \omega_n) + \liminf U(D(\mathfrak{I}) - \Delta_0, \omega_n) \\ &\geq U(D(\mathfrak{I}), \omega_0) = U(\mathfrak{I}, \omega_0). \end{aligned}$$

Hence the sign of equality holds throughout; in particular,

$$\limsup U(\Delta_0, \omega_n) = \liminf U(\Delta_0, \omega_n) = U(\Delta_0, \omega_0).$$

7. LEMMA. *Let $\phi_n(\Delta), \psi_n(\Delta)$ $(n = 0, 1, 2, \cdots)$ be a sequence of pairs of interval functions satisfying* II, (5.1), (5.2), (6.1)-(6.3). *Then*

$$\lim U(\mathfrak{I}, \phi_n) = U(\mathfrak{I}, \phi_0), \qquad \lim U(\mathfrak{I}, \psi_n) = U(\mathfrak{I}, \psi_0). \tag{7.1}$$

Proof. If Δ is any interval contained in $\mathfrak{I}$ and if $D(\Delta)$ is any subdivision, then (see II, §§2-4)

$$[\phi_n(D(\Delta))^2 + \psi_n(\Delta)^2]^{\frac{1}{2}} \leq [\phi_n(D(\Delta))^2 + \psi_n(D(\Delta))^2]^{\frac{1}{2}} \leq \omega_n(D(\Delta)) \leq U(\Delta, \omega_n).$$

Since this inequality holds for every choice of $D(\Delta)$, it follows that

$$U(\Delta, \phi_n)^2 \leq U(\Delta, \omega_n)^2 - \psi_n(\Delta)^2 \qquad (\Delta \subset \mathfrak{I}).$$

From II, (6.4), (5.2) follows

$$\{\limsup U(\Delta, \phi_n)\}^2 \leq U(\Delta, \omega_0)^2 - \psi_0(\Delta)^2 = U(\Delta, \omega_0)^2 - \omega_0(\Delta)^2 + \phi_0(\Delta)^2 \qquad (\Delta \subset \mathfrak{I}).$$

Since $\omega_0(\Delta) \leq U(\Delta, \omega_0)$, one concludes that

$$\limsup U(\Delta, \phi_n) \leq 2^{\frac{1}{2}}\, U(\Delta, \omega_0)^{\frac{1}{2}}\, \{U(\Delta, \omega_0) - \omega_0(\Delta)\}^{\frac{1}{2}} + \phi_0(\Delta) \qquad (\Delta \subset \mathfrak{I}).$$

Now let $D(\mathfrak{I})$ be any subdivision. Using this inequality, the additivity of the U-function, and the lemma of Schwarz, one obtains

$$\begin{aligned}\limsup U(\mathfrak{I}, \phi_n) &= \limsup U(D(\mathfrak{I}), \phi_n) \\ &\leq 2^{\frac{1}{2}}\, U(\mathfrak{I}, \omega_0)^{\frac{1}{2}} \{U(\mathfrak{I}, \omega_0) - \omega_0(D(\mathfrak{I}))\}^{\frac{1}{2}} + U(\mathfrak{I}, \phi_0).\end{aligned}$$

Since the first term in the right member of this inequality is arbitrarily small for the proper choice of $D(\mathfrak{I})$, it follows that

$$\limsup U(\mathfrak{I}, \phi_n) \leq U(\mathfrak{I}, \phi_0).$$

This relation, together with II, (5.4), implies the first relation in II, (7.1); the second relation follows similarly.

8. COROLLARY. *Let $\phi_n(\Delta), \psi_n(\Delta), \chi_n(\Delta)$ $(n = 0, 1, 2, \cdots)$ be a sequence of triples of interval functions. Set*

$$\lambda_n(\Delta) = [\psi_n(\Delta)^2 + \chi_n(\Delta)^2]^{\frac{1}{2}}, \qquad \mu_n(\Delta) = [\chi_n(\Delta)^2 + \phi_n(\Delta)^2]^{\frac{1}{2}},$$

$$\nu_n(\Delta) = [\phi_n(\Delta)^2 + \psi_n(\Delta)^2]^{\frac{1}{2}},$$

$$\begin{aligned}\omega_n(\Delta) &= [\phi_n(\Delta)^2 + \lambda_n(\Delta)^2]^{\frac{1}{2}} = [\psi_n(\Delta)^2 + \mu_n(\Delta)^2]^{\frac{1}{2}} = [\chi_n(\Delta)^2 + \nu_n(\Delta)^2]^{\frac{1}{2}} \\ &= [\phi_n(\Delta)^2 + \psi_n(\Delta)^2 + \chi_n(\Delta)^2]^{\frac{1}{2}} \qquad (\Delta \subset \mathfrak{I};\ n = 0, 1, 2, \cdots).\end{aligned}$$

Assume that the following conditions are satisfied.

(8.1) *The interval functions* $\phi_n(\Delta)$, $\psi_n(\Delta)$, $\chi_n(\Delta)$ *are non-negative for* $n = 0, 1, 2, \cdots$.

$$\begin{gathered}\liminf \phi_n(\Delta) \geq \phi_0(\Delta), \qquad \liminf \psi_n(\Delta) \geq \psi_0(\Delta), \\ \liminf \chi_n(\Delta) \geq \chi_0(\Delta)\end{gathered} \qquad (\Delta \subset \mathfrak{I}). \tag{8.2}$$

(8.3) *The interval functions* $\phi_n(\Delta)$, $\psi_n(\Delta)$, $\chi_n(\Delta)$ *increase by subdivision for* $n = 0, 1, 2, \cdots$.

(8.4) *The interval functions* $\phi_n(\Delta)$, $\psi_n(\Delta)$, $\chi_n(\Delta)$ *have U-functions for* $n = 0, 1, 2, \cdots$.

$$\lim U(\mathfrak{I}, \omega_n) = U(\mathfrak{I}, \omega_0). \tag{8.5}$$

Then

$$\begin{gathered}\lim U(\mathfrak{I}, \phi_n) = U(\mathfrak{I}, \phi_0), \qquad \lim U(\mathfrak{I}, \psi_n) = U(\mathfrak{I}, \psi_0), \\ \lim U(\mathfrak{I}, \chi_n) = U(\mathfrak{I}, \chi_0);\end{gathered} \tag{8.6}$$

$$\begin{gathered}\lim U(\mathfrak{I}, \lambda_n) = U(\mathfrak{I}, \lambda_0), \qquad \lim U(\mathfrak{I}, \mu_n) = U(\mathfrak{I}, \mu_0), \\ \lim U(\mathfrak{I}, \nu_n) = U(\mathfrak{I}, \nu_0).\end{gathered} \tag{8.7}$$

Proof. Observe that the pairs

$$[\phi_n(\Delta), \lambda_n(\Delta)], \qquad [\psi_n(\Delta), \mu_n(\Delta)], \qquad [\chi_n(\Delta), \nu_n(\Delta)] \qquad (n = 0, 1, 2, \cdots)$$

satisfy the conditions of the preceding lemma, and apply it.

In view of the lemma in II, §6, it is clear that the hypotheses of this corollary are also satisfied on any subinterval of $\mathfrak{I}$; thus relations (8.6) and (8.7) are valid on every subinterval of $\mathfrak{I}$.

9. In the sequel the following notation will be convenient. If $\mathfrak{x} = (X, Y, Z)$, $\mathfrak{u} = (U, V, W)$ are two triples, then set

$$\| \mathfrak{x} \| = [X^2 + Y^2 + Z^2]^{\frac{1}{2}}, \qquad \mathfrak{x} \pm \mathfrak{u} = (X \pm U, Y \pm V, Z \pm W),$$

$$\mathfrak{x} \times \mathfrak{u} = (YW - ZV, ZU - XW, XV - YU), \qquad \mathfrak{x} \cdot \mathfrak{u} = XU + YV + ZW.$$

10. Lemma. *Let* $\mathfrak{x}(u) = (X(u), Y(u), Z(u))$ *be a triple of real-valued functions defined and summable on the interval* Δ (see II, §1). *For any interval* δ *contained in* Δ, *define*

$$\phi(\delta) = \int_\delta X(u)\, du, \qquad \psi(\delta) = \int_\delta Y(u)\, du,$$

$$\chi(\Delta) = \int_\delta Z(u)\, du, \qquad \omega(\delta) = [\phi(\delta)^2 + \psi(\delta)^2 + \chi(\delta)^2]^{\frac{1}{2}}.$$

If $D_n(\Delta)$ be any sequence of subdivisions for which $\| D_n(\Delta) \|$ converges to zero, then it is true that (see II, §§2, 3)

$$\lim \omega(D_n(\Delta)) = \int_\Delta \| \mathfrak{x}(u) \| \, du. \tag{10.1}$$

$$U(\Delta, \omega) = \int_\Delta \| \mathfrak{x}(u) \| \, du. \tag{10.2}$$

Proof. If $D(\Delta)$ be any subdivision, it follows by a known inequality [3; VI] that

$$\omega(D(\Delta)) \leq \int_\Delta \| \mathfrak{x}(u) \| \, du.$$

Thus relation (10.2) will be established as soon as statement (10.1) is proved. In the special case when each of the functions $X(u)$, $Y(u)$, $Z(u)$ is a step-function —that is, is constant on the interior of each interval of some subdivision of Δ— the reader will verify (10.1). Consider the general case. Given a positive number ϵ, there exists a triple of step-functions $\mathfrak{u}(u) = (U(u), V(u), W(u))$ defined on Δ such that

$$\int_\Delta \| \mathfrak{x}(u) - \mathfrak{u}(u) \| \, du < \epsilon. \tag{10.3}$$

Denote by $\theta(\delta)$ the interval function defined in terms of $U(u)$, $V(u)$, $W(u)$ which corresponds to the function $\omega(\delta)$. Clearly

$$\left| \int_\Delta \| \mathfrak{x}(u) \| \, du - \int_\Delta \| \mathfrak{u}(u) \| \, du \right| \leq \int_\Delta \| \mathfrak{x}(u) - \mathfrak{u}(u) \| \, du < \epsilon. \tag{10.4}$$

For any interval δ contained in Δ, it follows that

$$| \omega(\delta) - \theta(\delta) | \leq \int_\delta \| \mathfrak{x}(u) - \mathfrak{u}(u) \| \, du;$$

hence, from inequality (10.3), one obtains

$$| \omega(D_n(\Delta)) - \theta(D_n(\Delta)) | \leq \int_\Delta \| \mathfrak{x}(u) - \mathfrak{u}(u) \| \, du < \epsilon. \tag{10.5}$$

Inequalities (10.4) and (10.5), together with the fact that (10.1) is verified for the triple $\mathfrak{u}(u)$, imply that

$$\int_{\Delta} || \mathfrak{x}(u) || \, du - 2\epsilon \leq \liminf \omega(D_n(\Delta))$$

$$\leq \limsup \omega(D_n(\Delta)) \leq \int_{\Delta} || \mathfrak{x}(u) || + 2\epsilon.$$

Since ϵ is an arbitrary positive number, relation (10.1) is verified.

11. Lemma. *Let E be any measurable set in n-dimensional Euclidean space. If*

$$\mathfrak{x}(u) = (X(u), Y(u), Z(u)), \qquad \mathfrak{u}(u) = (U(u), V(u), W(u)) \qquad (u \in E)$$

be two triples of real-valued functions defined and measurable on E then (see II, §9)

$$\left(\int_E || \mathfrak{x}(u) \times \mathfrak{u}(u) || \, du \right)^2 \leq \left(\int_E || \mathfrak{x}(u) || \, || \mathfrak{u}(u) || \, du \right)^2 - \left(\int_E \mathfrak{x}(u) \cdot \mathfrak{u}(u) \, du \right)^2 \tag{11.1}$$

as soon as the integrals involved exist.

Proof. From the identity

$$(|| \mathfrak{x}(u) || \, || \mathfrak{u}(u) || \,)^2 = || \mathfrak{x}(u) \times \mathfrak{u}(u) ||^2 + (\mathfrak{x}(u) \cdot \mathfrak{u}(u))^2 \qquad (u \in E)$$

there follows by integration and a known inequality [3; VI]

$$\int_E || \mathfrak{x}(u) || \, || \mathfrak{u}(u) || \, du = \int_E \{ || \mathfrak{x}(u) \times \mathfrak{u}(u) ||^2 + (\mathfrak{x}(u) \cdot \mathfrak{u}(u))^2 \}^{\frac{1}{2}} du$$

$$\geq \left\{ \left(\int_E || \mathfrak{x}(u) \times \mathfrak{u}(u) || \, du \right)^2 + \left(\int_E \mathfrak{x}(u) \cdot \mathfrak{u}(u) \, du \right)^2 \right\}^{\frac{1}{2}},$$

whence (11.1) follows.

12. Corollary. *For $\mathfrak{u}(u) = \mathfrak{a} =$ constant, relation* II, (11.1) *becomes*

$$\left(\int_E || \mathfrak{x}(u) \times \mathfrak{a} || \, du \right)^2 \leq || \mathfrak{a} ||^2 \left(\int_E || \mathfrak{x}(u) || \, du \right)^2 - \left(\mathfrak{a} \cdot \int_E \mathfrak{x}(u) \, du \right)^2. \tag{12.1}$$

13. COROLLARY. *Assume that* $| E | > 0$, *and choose*

$$\mathfrak{a} = \left(| E |^{-1} \int_E X(u)\, du,\ | E |^{-1} \int_E Y(u)\, du,\ | E |^{-1} \int_E Z(u)\, du \right). \tag{13.1}$$

Then II, (12.1) becomes

$$\left(\int_E || \mathfrak{x}(u) \times \mathfrak{a} ||\, du \right)^2 \leq || \mathfrak{a} ||^2 \left\{ \left(\int_E || \mathfrak{x}(u) ||\, du \right)^2 - | E |^2 || \mathfrak{a} ||^2 \right\}. \tag{13.2}$$

14. COROLLARY. *If* $\mathfrak{x}(u) = (X(u), Y(u), 1)$, $u \in E$, *and* $\mathfrak{a}$ *is defined by* II, (13.1), *then from* II (13.2) *follows*

$$\left(\int_E || \mathfrak{x}(u) - \mathfrak{a} ||\, du \right)^2 \leq \left(\int_E || \mathfrak{x}(u) \times \mathfrak{a} ||\, du \right)^2 \leq || \mathfrak{a} ||^2 \left\{ \left(\int_E || \mathfrak{x}(u) ||\, du \right)^2 - | E |^2 || \mathfrak{a} ||^2 \right\}, \tag{14.1}$$

or

$$\int_E || \mathfrak{x}(u) - \mathfrak{a} ||\, du \leq 2^{\frac{1}{2}} || \mathfrak{a} || \left(\int_E || \mathfrak{x}(u) ||\, du \right)^{\frac{1}{2}} \times \left\{ \int_E || \mathfrak{x}(u) ||\, du - \int_E || \mathfrak{a} ||\, du \right\}^{\frac{1}{2}}, \tag{14.2}$$

since

$$| E |\, || \mathfrak{a} || = \int_E || \mathfrak{a} ||\, du \leq \int_E || \mathfrak{x}(u) ||\, du.$$

15. The reader will observe that the corollary in II, §14 contains all that we shall need for our second proof of strong convergence for non-parametric representations of surfaces (see III, §§6-13). Its usefulness lies in the fact that, for vectors having the special form given in II, §14, we have an estimate for the difference $|| \mathfrak{x}(u) - \mathfrak{a} ||$, given in II, (14.1), (14.2).

Furthermore, the corollary in II, §14 contains as a special case the fundamental inequality used by Adams and Lewy in proving their result on strong convergence

for curves (see I, §§1-3). Using the notation of I, §1, we may state their inequality as follows. If $f(x)$ is any function of bounded variation on $[a, b]$, then

$$[L(f)]^2 - (b-a)^2(1+m^2) \geq [T(f-mx)]^2/(1+m^2), \qquad m = [f(b)-f(a)]/(b-a). \tag{15.1}$$

Now if $f(x)$ be absolutely continuous on $[a, b]$, then (15.1) becomes (see I, §6)

$$\left[\int_a^b \{f'(x)^2+1\}^{\frac{1}{2}}\,dx\right]^2 - (b-a)^2(1+m^2) \geq \left[\int_a^b |f'(x)-m|\,dx\right]^2/(1+m^2), \qquad m = \int_a^b f'(x)\,dx/(b-a). \tag{15.2}$$

Put $E = [a, b]$, $\mathfrak{a} = (0, m, 1)$, $\mathfrak{x} = (0, f', 1)$ in II, (14.1), and (15.2) results. Inequality (15.1) follows at once from (15.2) by approximating $f(x)$ on $[a, b]$ by a sequence of absolutely continuous functions, say inscribed polygons, which converge in length to $f(x)$ on $[a, b]$, see [2].

16. The lemmas in the sequel are not new, but since we are aware of no convenient reference, we present them for the convenience of the reader.

(16.1) Let E be a measurable subset of Euclidean n-space, having finite measure.
(16.2) Let $Y_n(u)$ $(n = 1, 2, \cdots)$ be a sequence of real-valued functions, each defined and measurable on E.

The following notation will be convenient for the sequel. Given a positive number ϵ, denote by $e_n(+\epsilon)$, $e_n(-\epsilon)$, the set of those points contained in E for which $Y_n(u) > +\epsilon$, $Y_n(u) < -\epsilon$ respectively. Set

$$E_m(+\epsilon) = \sum_{n \leq m} e_n(+\epsilon), \qquad E_m(-\epsilon) = \sum_{n \leq m} e_n(-\epsilon),$$

$$e_n(\epsilon) = e_n(+\epsilon) + e_n(-\epsilon).$$

Observe that each of these sets is measurable. The sequence $Y_n(u)$ is said to converge to zero in measure on E—briefly, $Y_n - m \to 0$ on E—provided, for every positive number ϵ, it is true that the measures of the sets $e_n(\epsilon)$ converge to zero. A sequence of real-valued functions $X_n(u)$, defined and measurable on the set E, is said to converge in measure to the real-valued, measurable function $X_0(u)$ on E—briefly, $X_n - m \to X_0$ on E—provided $(X_n - X_0) - m \to 0$ on E.

17. Lemma. *Let $Y_n(u)$ $(n = 1, 2, \cdots)$ and E satisfy* II, (16.1), (16.2) *and*

(17.1) $\liminf Y_n(u) \geq 0$ *almost everywhere on E.*

Then

(17.2) $\lim |E_n(-\epsilon)| = 0$ *for* $\epsilon > 0$.

Proof. This follows immediately from the facts that $E_m(-\epsilon) \supset E_{m+1}(-\epsilon)$ and $\prod E_m(-\epsilon)$ is a set of measure zero.

COROLLARY. *If condition* (17.1) *is replaced by*

(17.3) $\limsup Y_n(u) \leq 0$ *almost everywhere on* E,

then

(17.4) $\lim | E_n(+\epsilon) | = 0$ *for* $\epsilon > 0$.

These results imply, in particular, the

COROLLARY. *If* $\lim Y_n(u) = 0$ *almost everywhere on* E, *then* $Y_n - m \to 0$ *on* E.

18. The converse of the preceding corollary is generally false, but we have the following

LEMMA. *Let* $Y_n(u)$ $(n = 1, 2, \cdots)$ *and* E *satisfy* II, (16.1), (16.2) *and*

(18.1) $Y_n - m \to 0$ *on* E.

Then there exists a subsequence of functions $Y_{n_k}(u)$ *and a sequence of sets* E_k *satisfying the following conditions.*

(18.2) $E_k \supset E_{k+1}$, $\lim | E_k | = 0$.
(18.3) $| Y_{n_k}(u) | \leq k^{-1}$ *for* $u \in E - E_k$.
(18.4) $\lim Y_{n_k}(u) = 0$ *almost everywhere on* E.

Proof. If $| E | = 0$, the lemma is trivial; assume $| E | > 0$. Statement (18.4) follows at once from (18.2) and (18.3); the latter may be proved by the diagonal process. First, it follows from condition (18.1) (see II, §16) that there exists a subsequence of functions $Y_{1j}(u)$, and a sequence of sets e_{1j} contained in E for which

$$| Y_{1j}(u) | \leq j^{-1}, \qquad u \in E - e_{1j}, \qquad | e_{1j} | < | E | \cdot 2^{-j-1}.$$

Set $E_1 = \sum e_{1j}$. Then $| E_1 | < | E | \cdot 2^{-1}$, and clearly $Y_{1j} - m \to 0$ on E_1. Repeating the above argument for the sequence of functions $Y_{1j}(u)$ and the set E_1, we obtain a subsequence of functions $Y_{2j}(u)$ and a sequence of sets e_{2j} contained in E_1 for which

$$| Y_{2j}(u) | \leq j^{-1}, \qquad u \in E - e_{2j}, \qquad | e_{2j} | < | E_1 | \cdot 2^{-j-1}.$$

Set $E_2 = \sum e_{2j}$. Then $| E_2 | < | E_1 | \cdot 2^{-1} < | E | \cdot 2^{-2}$. Proceed in this fashion, and choose $Y_{n_k}(u)$ to be the function $Y_{kk}(u)$ defined in the k-th step. The reader will verify that these functions $Y_{n_k}(u)$, together with the sets E_k defined in this process, satisfy conditions (18.2) and (18.3).

19. LEMMA. *Let* $Y_n(u)$ $(n = 1, 2, \cdots)$ *and* E *satisfy* II, (16.1), (16.2), (17.1) *and*

(19.1) *each of the functions* $Y_n(u)$ $(n = 1, 2, \cdots)$ *is summable on* E;

(19.2) *if* e *be any measurable subset of* E, *then* $\lim \int_e Y_n(u) = 0$.

Then

(19.3) $Y_n - m \to 0$ *on* E.

Proof. Using the notation of II, §16, we clearly have, for $n \geq m$, $\epsilon > 0$, $\tau > 0$,

$$\int_E Y_n(u)\,du = \int_{E_m(-\tau)} Y_n(u)\,du + \int_{e_n(+\epsilon)-E_m(-\tau)} Y_n(u)\,du + \int_{E-(E_m(-\tau)+e_n(+\epsilon))} Y_n(u)\,du$$

$$\geq \int_{E_m(-\tau)} Y_n(u)\,du + \epsilon \mid e_n(+\epsilon) \mid - \epsilon \mid E_m(-\tau) \mid - \tau \mid E \mid.$$

From condition (19.2), it follows that, for m fixed, $\epsilon > 0$, $\tau > 0$,

$$\limsup \mid e_n(+\epsilon) \mid \leq \tau \mid E \mid \cdot \epsilon^{-1} + \mid E_m(-\tau) \mid.$$

From the lemma in II, §17 and the fact that τ is an arbitrary positive number, we obtain

$$\lim \mid e_n(+\epsilon) \mid = 0 \qquad (\epsilon > 0). \tag{19.4}$$

From the lemma in II, §17, it follows at once that

$$\lim \mid e_n(-\epsilon) \mid = 0 \qquad (\epsilon > 0). \tag{19.5}$$

Relations (19.4) and (19.5) imply the statement in (19.3) (see II, §16).

20. LEMMA. *Let there be given*

(20.1) *a measurable subset* E *of Euclidean* n*-space, having finite measure*;
(20.2) *a sequence of real-valued functions* $X_n(u)$ $(n = 0, 1, 2, \cdots)$, *defined on* E.

Assume that the following conditions are satisfied.

(20.3) *Each of the functions* $X_n(u)$ *for* $n = 0, 1, 2, \cdots$ *is measurable and summable on* E.
(20.4) *Each of the functions* $X_n(u)$ *for* $n = 1, 2, \cdots$ *is non-negative on* E.
(20.5) $\liminf X_n(u) \geq X_0(u)$ *almost everywhere on* E.

$$\limsup \int_E X_n(u)\,du \leq \int_E X_0(u)\,du. \tag{20.6}$$

Then

(20.7) *if e be a measurable subset of E, it is true that* $\lim \int_e X_n(u)du = \int_e X_0(u)du$;

(20.8) *there exists a subsequence of functions $X_{n_k}(u)$ such that* $\lim X_{n_k}(u) = X_0(u)$ *almost everywhere on E.*

Proof. From the lemma of Fatou [13; I], it follows that, for any measurable subset e of E,

$$\liminf \int_e X_n(u)\,du \geq \int_e \{\liminf X_n(u)\}\,du \geq \int_e X_0(u)\,du.$$

Thus, using condition (20.6), we have

$$\begin{aligned}
\int_E X_0(u)\,du &\geq \limsup \int_E X_n(u)\,du \\
&\geq \limsup \int_e X_n(u)\,du + \liminf \int_{E-e} X_n(u)\,du \\
&\geq \liminf \int_e X_n(u)\,du + \liminf \int_{E-e} X_n(u)\,du \\
&\geq \int_e X_0(u)\,du + \int_{E-e} X_0(u)\,du = \int_E X_0(u)\,du.
\end{aligned}$$

Since the sign of equality must hold throughout, statement (20.7) is verified. It is now clear that the functions

$$Y_n(u) \equiv X_n(u) - X_0(u) \qquad (n = 1, 2, 3, \cdots)$$

satisfy the hypotheses of the lemma in II, §19, and thus assertion (20.8) follows at once from the lemma in II, §18.

21. LEMMA. *Let $Y_n(u)$ $(n = 1, 2, \cdots)$ and E satisfy* II, (16.1), (16.2), (18.1) *and let there be given a sequence of real-valued functions $Z_n(u)$ $(n = 0, 1, 2, \cdots)$ each defined, non-negative, measurable, and summable on E, for which the following conditions are satisfied.*

(21.1) $|\, Y_n(u)\,| \leq Z_n(u)$ $(n = 1, 2, \cdots)$ *almost everywhere on E.*

(21.2) *On any measurable subset e of E,* $\limsup \int_e Z_n(u)du \leq \int_e Z_0(u)du$.

Then

$$\lim \int_E | Y_n(u) | du = 0. \tag{21.3}$$

Proof. Condition (21.1) implies that the $Y_n(u)$ $(n = 1, 2, \cdots)$ are summable. The hypotheses of the lemma in II, §18 being satisfied, its conclusions follow; using the notations of that lemma, and condition (21.1) we find for $m \geq k$,

$$\int_E | Y_{n_m}(u) | du = \int_{E-E_k} + \int_{E_k} | Y_{n_m}(u) | du \leq | E | \cdot m^{-1} + \int_{E_k} Z_{n_m}(u) \, du.$$

In view of condition (21.2), we have, for fixed k,

$$\limsup \int_E | Y_{n_m}(u) | du \leq \int_{E_k} Z_0(u) \, du.$$

From II, (18.2) and the absolute continuity of the integral, it follows that

$$\lim \int_E | Y_{n_m}(u) | du = 0.$$

But since any infinite subsequence of the functions $Y_n(u)$ $(n = 1, 2, \cdots)$ satisfies the same hypotheses as the sequence $Y_n(u)$ $(n = 1, 2, \cdots)$, relation (21.3) follows.

CHAPTER III

Proofs

1. First, we prove the theorem stated in I, §8. Using the notation introduced in stating that theorem, we define, for any interval $\Delta = [u', u'']$ contained in $\mathfrak{I} = [a, b]$, and for $n = 0, 1, 2, \cdots$,

$$\begin{gathered} \phi_n(\Delta) = | x_n(u') - x_n(u'') |, \qquad \psi_n(\Delta) = | y_n(u') - y_n(u'') |, \\ \chi_n(\Delta) = |z_n(u') - z_n(u'')|. \end{gathered} \tag{1.1}$$

From the definitions of total variation, length, and U-function (see II, §3), it is clear that (see I, §7 and II, §8)

$$\begin{aligned} &U(\mathfrak{I}, \phi_n) = T(x_n), \qquad U(\mathfrak{I}, \psi_n) = T(y_n), \qquad U(\mathfrak{I}, \chi_n) = T(z_n), \\ &U(\mathfrak{I}, \lambda_n) = L(x\text{-}\mathfrak{x}_n), \qquad U(\mathfrak{I}, \mu_n) = L(y\text{-}\mathfrak{x}_n), \\ &U(\mathfrak{I}, \nu_n) = L(z\text{-}\mathfrak{x}_n), \qquad U(\mathfrak{I}, \omega_n) = L(\mathfrak{x}_n) \qquad (n = 0, 1, 2, \cdots). \end{aligned}$$

Using the hypothesis made in I, (8.1), (8.2), the reader will find that the interval functions in (1.1) satisfy the assumptions made for the corollary in II, §8. In view of the above relations and the definitions in I, §7, the assertions in I, (8.1), (8.2) now follow from II, (8.6), (8.7). To establish I, (8.3), we give an example. Choose $\mathfrak{I} = [0, 1]$ and define $p_n(u)$ for $n = 1, 2, \cdots$ to be the continuous, piecewise linear function satisfying $p_n(0) = 0$; $p_n(1) = 1$; $p_n(u)$ has slope $3/2$ on intervals $[(2k - 2)\cdot 2^{-n}, (2k - 1)\, 2^{-n}]$ and slope $\frac{1}{2}$ on intervals $[(2k - 1)\, 2^{-n}, (2k)2^{-n}]$ for $k = 1, \cdots, 2^{n-1}$. (These functions are similar to those defined in [1] for another purpose; the functions used by Adams and Clarkson would also serve here.) Define

$$\mathfrak{x}_n(u) = (p_n(u), p_n(u), p_n(u)), \qquad \mathfrak{x}_0(u) = (u, u, u) \qquad (u \in \mathfrak{I}).$$

Clearly the $p_n(u)$ are absolutely continuous, converge uniformly on $\mathfrak{I}$ to u, and (see I, §6)

$$L(\mathfrak{x}_n) = 3^{\frac{1}{2}} \int_0^1 | p_n'(u) | \, du = 3^{\frac{1}{2}} \qquad (n = 1, 2, \cdots);$$

$$L(\mathfrak{x}_0) = 3^{\frac{1}{2}}.$$

Thus $\mathfrak{x}_n - l \to \mathfrak{x}_0$ on $\mathfrak{I}$, but

$$T(p_n(u) - u) = \int_0^1 | p_n'(u) - 1 | \, du = 2^{-1} \quad (n = 1, 2, \cdots).$$

Therefore, $\mathfrak{x}_n$ does not converge strongly in variation to $\mathfrak{x}_0$ on $\mathfrak{I}$.

2. To verify I, (8.4), it suffices to prove a sharper result. In stating it, we use the notation of I, §1.

LEMMA. *If a sequence of functions $f_n(x)$ $(n = 0, 1, 2, \cdots)$ satisfies conditions* I, (1.1)-(1.3), *and if, moreover, each of the functions $f_n(x)$ is absolutely continuous on $[a, b]$ for $n = 1, 2, \cdots$, then a necessary condition that*

$$\lim \int_a^b | f_n'(x) - f_0'(x) | \, dx = 0$$

is that $f_0(x)$ also be absolutely continuous on $[a, b]$.

Proof. From I, (1.3), it follows that the derivatives $f_n'(x)$ $(n = 0, 1, 2, \cdots)$ exist almost everywhere on $[a, b]$ and are summable. From the following inequality, valid almost everywhere on $[a, b]$,

$$| f_n'(x) - f_0'(x) | \geq [f_n'(x)^2 + 1]^{\frac{1}{2}} - [f_0'(x)^2 + 1]^{\frac{1}{2}},$$

we obtain by integration and the fact that the $f_n(x)$ are absolutely continuous for $n = 1, 2, \cdots$ (see I, §6)

$$\int_a^b | f_n'(x) - f_0'(x) | \, dx \geq L(f_n) - \int_a^b [f_0'(x)^2 + 1]^{\frac{1}{2}} \, dx.$$

Since $\liminf L(f_n) \geq L(f_0)$, it follows that

$$\liminf \int_a^b | f_n'(x) - f_0'(x) | \, dx \geq L(f_0) - \int_a^b [f_0'(x)^2 + 1]^{\frac{1}{2}} \, dx.$$

Now the right member of this inequality is greater than zero unless $f_0(x)$ is absolutely continuous; thus the lemma is proved.

Since we have

$$T(f_n - f_0) \geq \int_a^b | f_n'(x) - f_0'(x) | \, dx,$$

it follows at once that if $f_n - sv \to f_0$ on $[a, b]$, then $f_0(x)$ is absolutely continuous on $[a, b]$. The statement in I, (8.4) is now immediate (see I, (7.5)).

3. Next, we prove the theorem stated in I, §10. Introduce parametric representations for the curves $\mathfrak{y} = \mathfrak{y}_n(x)$ $(x \in [a, b]; n = 0, 1, 2, \cdots)$ as follows. Let $\mathfrak{I}$ denote the interval $[a, b]$ on the u-axis, and set $x = u$, $u \in \mathfrak{I}$. Define

$$\mathfrak{x}_n(u) = (u, y_n(u), z_n(u)) \qquad (u \in \mathfrak{I}; n = 0, 1, 2, \cdots).$$

Clearly the statements $\mathfrak{y}_n - v \to \mathfrak{y}_0$, $\mathfrak{y}_n - sv \to \mathfrak{y}_0$, $\mathfrak{y}_n - l \to \mathfrak{y}_0$, $y_n - l \to y_0$, $z_n - l \to z_0$ on $[a, b]$ are equivalent to the statements $\mathfrak{x}_n - v \to \mathfrak{x}_0$, $\mathfrak{x}_n - sv \to \mathfrak{x}_0$, $\mathfrak{x}_n - l \to \mathfrak{x}_0$, $z\text{-}\mathfrak{x}_n - l \to z\text{-}\mathfrak{x}_0$, $y\text{-}\mathfrak{x}_n - l \to y\text{-}\mathfrak{x}_0$ on $\mathfrak{I}$, respectively. From I, (8.1), (8.2), (8.4), the statements in I, (10.1), (10.2), (10.4) follow at once. Since, by I, (10.1), the fact that $\mathfrak{y}_n - l \to \mathfrak{y}_0$ on $[a, b]$ implies that $y_n - l \to y_0$ and $z_n - l \to z_0$ on $[a, b]$, statement I, (10.3) is a consequence of the result of Adams and Lewy cited in I, (3.2).

4. We turn to prove the theorem stated in I, §19. Using the notation introduced in stating that theorem, we define, for every interval $\Delta = [x', x''; y', y']$ contained in $\mathfrak{I} = [a, b; c, d]$, and for $n = 0, 1, 2, \cdots$ (see II, §8),

$$\phi_n(\Delta) = \int_{y'}^{y''} | f_n(x', y) - f_n(x'', y) | \, dy,$$

(4.1). $$\psi_n(\Delta) = \int_{x'}^{x''} | f_n(x, y') - f_n(x, y'') | \, dx,$$

$$\chi_n(\Delta) = (x'' - x')(y'' - y') = |\Delta|.$$

(These are the so-called expressions of Geocze [5], [6].) Interval functions $\lambda_n(\Delta)$, $\mu_n(\Delta)$, $\nu_n(\Delta)$, $\omega_n(\Delta)$ are defined as in II, §8; we want to verify that these interval functions satisfy the hypotheses of that corollary. Clearly condition II, (8.1) is satisfied. In view of I, (8.2), it follows that II, (8.2) is fulfilled; in fact, $\lim \phi_n(\Delta) = \phi_0(\Delta)$, $\lim \psi_n(\Delta) = \psi_0(\Delta)$, $\lim \chi_n(\Delta) = \chi_0(\Delta)$. Evidently the

interval functions in (4.1) satisfy II, (8.3); in fact, $\chi_n(\Delta)$ is additive. We assert that (see I, §§12-14)

$$(4.2)\qquad \begin{aligned} &U(\mathfrak{I}, \phi_n) = T_x(f_n), \qquad U(\mathfrak{I}, \psi_n) = T_y(f_n), \qquad U(\mathfrak{I}, \chi_n) = |\,\mathfrak{I}\,|, \\ &U(\mathfrak{I}, \lambda_n) = S_y(f_n), \qquad U(\mathfrak{I}, \mu_n) = S_x(f_n) \qquad (n = 0, 1, 2, \cdots); \end{aligned}$$

so that, in particular, condition II, (8.4) is fulfilled (see I, §§17, 18). The third relation in this set is obvious; the others may be established by a method which we now illustrate in proving the fourth relation in this set. For simplicity, we shall omit the subscript n. Since $\lambda(\Delta)$ increases by subdivision, it is sufficient, in computing $U(\mathfrak{I}, \lambda)$, to consider only subdivisions of $\mathfrak{I}$ (see II, §1) in which all the lines of subdivision extend from boundary to boundary in $\mathfrak{I}$. Such a subdivision $D(\mathfrak{I})$ may be specified by giving two linear subdivisions

$$D([a, b]): a = x_0 < \cdots < x_i < \cdots < x_m = b;$$

$$D([c, d]): c = y_0 < \cdots < y_j < \cdots < y_n = d.$$

From elementary inequalities, we have (see I, §14)

$$\begin{aligned} \lambda(D(\mathfrak{I})) = \sum_{i=1}^{m}\sum_{j=1}^{n}\Bigg[\left\{\int_{x_{i-1}}^{x_i} |\, f(x, y_{j-1}) - f(x, y_j)\,|\, dx\right\}^2 \\ + \left\{\int_{x_{i-1}}^{x_i} |\, y_{j-1} - y_j\,|\, dy\right\}^2\Bigg]^{\frac{1}{2}} \leq \sum_{i=1}^{m}\int_{x_{i-1}}^{x_i}\sum_{j=1}^{n}[\{f(x, y_{j-1}) - f(x, y_j)\}^2 \\ + \{y_{j-1} - y_j\}^2]^{\frac{1}{2}}\, dx \leq \int_a^b L_y(x, f)\, dx = S_y(f). \end{aligned}$$

Thus (see II, §3)

$$(4.3)\qquad U(\mathfrak{I}, \lambda) \leq S_y(f).$$

To get the opposite inequality, we note that from II, §10 follows

$$\begin{aligned} \lim_{\|D([a,b])\|\to 0} \sum_{i=1}^{m}\Bigg[\left\{\int_{x_{i-1}}^{x_i} |\, f(x, y_{j-1}) - f(x, y_j)\,|\, dx\right\}^2 + \left\{\int_{x_{i-1}}^{x_i} |\, y_{j-1} - y_j\, dx\right\}^2\Bigg]^{\frac{1}{2}} \\ = \int_a^b [\{f(x, y_{j-1}) - f(x, y_j)\}^2 + \{y_{j-1} - y_j\}^2]^{\frac{1}{2}}\, dx. \end{aligned}$$

Hence, for all subdivisions $D([c, d])$, it is true that

$$U(\mathfrak{I}, \lambda) \geq \int_a^b \sum_{j=1}^{n}[\{f(x, y_{j-1}) - f(x, y_j)\}^2 + \{y_{j-1} - y_j\}^2]^{\frac{1}{2}}\, dx.$$

Define

$$g(x; D([c, d])) = \sum_{j=1}^{n}[\{f(x, y_{j-1}) - f(x, y_j)\}^2 + \{y_{j-1} - y_j\}^2]^{\frac{1}{2}}\, dx \quad (x \in [a, b]).$$

Then $g(x; D([c, d]))$ is a non-negative, continuous function of x on $[a, b]$, and

$$\lim_{\|D([c,d])\|\to 0} g(x; D([c, d])) = L_v(x, f).$$

Applying the lemma of Fatou [13; I], we obtain

$$U(\mathfrak{I}, \lambda) \geq \int_a^b L_v(x, f)\, dx = S_v(f). \tag{4.4}$$

Inequalities (4.3) and (4.4) imply that $U(\mathfrak{I}, \lambda) = S_v(f)$, as asserted. Now Radó [5], [6] has shown that

$$U(\mathfrak{I}, \omega_n) = A(f_n) \qquad (n = 0, 1, 2, \cdots).$$

It follows that condition II, (8.5) is fulfilled whenever $f_n - a \to f_0$ on $[a, b; c, d]$ (see I, (18.9)). Thus, in view of relations (4.2), the statements in I, (19.1), (19.2) are immediate consequences of the corollary in II, §8.

5. The fact that $f_n - a \to f_0$ on $[a, b; c, d]$ does not generally imply that $f_n - sv \to f_0$ on $[a, b; c, d]$ follows at once from the corresponding result of Adams and Lewy for curves (see I, (3.2)). But if $f_n - a \to f_0$ on $[a, b; c, d]$, then by I, (19.1), (19.2), we have, in particular,

$$\begin{aligned} \lim \int_c^d L_x(y, f_n)\, dy &= \int_c^d L_x(y, f_0)\, dy, \\ \lim \int_c^d V_x(y, f_n)\, dy &= \int_c^d V_x(y, f_0)\, dy. \end{aligned} \tag{5.1}$$

In view of I, (18.1), (18.2) it is clear that, for every $y = \eta$ contained in $[c, d]$, the functions $f_n(x, \eta)$ $(n = 0, 1, 2, \cdots)$ regarded as functions of x alone on the interval $[a, b]$ satisfy the conditions I, (1.1), (1.2). In particular, then (see I, §6), we have

$$\liminf L_x(y, f_n) \geq L_x(y, f_0) \qquad (y \in [c, d]).$$

The reader now observes that the functions $L_x(y, f_n)$ $(n = 0, 1, 2, \cdots)$ satisfy the hypotheses of the lemma in II, §20 on the interval $[c, d]$. Consequently there exists a subsequence of functions $L_x(y, f_{n_k})$ $(k = 1, 2, \cdots)$ such that

$$\lim L_x(y, f_{n_k}) = L_x(y, f_0) \qquad \text{for almost every } y \in [c, d]. \tag{5.2}$$

In view of I, (18.3), it follows that, for almost every $y = \eta$ contained in $[c, d]$, the functions $f_n(x, \eta)$ $(n = 0, 1, 2, \cdots)$ regarded as functions of x alone on the interval $[a, b]$, satisfy the condition I, (1.3). Hence, from relation (5.2), we conclude that (see I, §1)

$$f_{n_k}(x, \eta) - l \to f_0(x, \eta) \qquad \text{for almost every } y = \eta \in [c, d].$$

From the results of Adams and Lewy cited in I, (3.2), it follows that (see I, §§12, 13)

$$\lim V_x(y, f_{n_k} - f_0) = 0 \qquad \text{for almost every } y \in E(y, f_0). \tag{5.3}$$

Clearly

$$V_x(y, f_{n_k} - f_0) \leq V_x(y, f_{n_k}) + V_x(y, f_0) \qquad (y \in [c, d]). \tag{5.4}$$

Using relations (5.1)-(5.4) (see II, §17), the reader will find that the functions

$$V_x(y, f_{n_k} - f_0), \qquad V_x(y, f_{n_k}) + V_x(y, f_0) \qquad (k = 1, 2, 3, \cdots)$$

satisfy the assumptions of the lemma in II, §21 on the set $E(y, f_0)$. Consequently,

$$\lim \int_{E(y, f_0)} V_x(y, f_{n_k} - f_0)\, dy = 0.$$

But since this same reasoning may be applied to any infinite subsequence of the functions $f_n(x, y)$ $(n = 1, 2, \cdots)$, we conclude that

$$\lim \int_{E(y, f_0)} V_x(y, f_n - f_0)\, dy = 0.$$

This is the first relation in I, (19.3); the second follows at once by interchanging the rôles of x and y.

Since I, (19.4) has been verified in I, §20, the theorem in I, §19 is now established.

6. In the following sections, we give a second proof for the fact that if $f_n - a \to f_0$ on $[a, b; c, d]$, and if $f_0(x, y)$ is ACT on $[a, b; c, d]$, then $f_n - sv \to f_0$ on $[a, b; c, d]$ (see I, (19.3)). This proof is independent of the results of Adams and Lewy, but parallels closely their proof for the corresponding theorem on curves. It is based upon the inequality in II, (14.2). In addition to the notation introduced in stating the theorem in I, §19, we shall find the following notation convenient.

$$\pi(f) = |\, \Im \,|^{-1} \int_c^d \{f(b, y) - f(a, y)\}\, dy,$$

$$\rho(f) = |\, \Im \,|^{-1} \int_a^b \{f(x, d) - f(x, c)\}\, dx, \tag{6.1}$$

$$\sigma(f) = [\pi(f)^2 + \rho(f)^2 + 1]^{\frac{1}{2}},$$

$$\omega(f) = |\Im|\sigma(f) \qquad (\Im = [a, b; c, d]).$$

7. First, suppose that $f(x, y)$ is ACT on $\mathfrak{I}$; then (see I, §17)

$$\pi(f) = |\mathfrak{I}|^{-1} \iint_{\mathfrak{I}} p(x, y)\, dx\, dy, \qquad \rho(f) = |\mathfrak{I}|^{-1} \iint_{\mathfrak{I}} q(x, y)\, dx\, dy;$$

$$A(f) = \iint_{\mathfrak{I}} [p(x, y)^2 + q(x, y)^2 + 1]^{\frac{1}{2}}\, dx\, dy; \tag{7.1}$$

$$T_x(f - \pi(f)x) = \iint_{\mathfrak{I}} |\, p(x, y) - \pi(f)\,|\, dx\, dy.$$

If we identify

$$E \equiv \mathfrak{I}, \qquad u \equiv (x, y), \qquad X(u) \equiv p(x, y), \qquad Y(u) \equiv q(x, y),$$

then clearly the hypotheses of the corollary in II, §14 are satisfied. Thus from II, (14.2) follows

$$T_x(f - \pi(f)x) \leq 2^{\frac{1}{2}}\sigma(f)\, A(f)^{\frac{1}{2}}\{A(f) - \omega(f)\}^{\frac{1}{2}}. \tag{7.2}$$

8. Next, assume that $f_n - a \to f_0$ on $\mathfrak{I}$ (see I, (18.9)) so that, in particular, $\lim A(f_n) = A(f_0)$. Now clearly (see I, (18.2), III, (6.1))

$$\begin{aligned} &\lim \pi(f_n) = \pi(f_0), \qquad \lim \rho(f_n) = \rho(f_0), \\ &\lim \sigma(f_n) = \sigma(f_0), \qquad \lim \omega(f_n) = \omega(f_0). \end{aligned} \tag{8.1}$$

Thus it is clear that $f_n(x, y) - \pi(f_n)x$ converges uniformly on $\mathfrak{I}$ to $f_0(x, y) - \pi(f_0)x$; hence (see I, (17.4)) $\liminf T_x(f_n - \pi(f_n)x) \geq T_x(f_0 - \pi(f_0)x)$. Assume further that each $f_n(x, y)$ for $n = 1, 2, \cdots$ satisfies III, (7.2). Using the preceding relations, we find that $f_0(x, y)$ also satisfies III, (7.2)—in fact,

$$\begin{aligned} T_x(f_0 - \pi(f_0)x) &\leq \limsup T_x(f_n - \pi(f_n)x) \\ &\leq 2^{\frac{1}{2}}\sigma(f_0)\, A(f_0)^{\frac{1}{2}}\{A(f_0) - \omega(f_0)\}^{\frac{1}{2}}. \end{aligned} \tag{8.2}$$

9. Assume now that $f_0(x, y)$ is merely BVT on $\mathfrak{I}$; from the definition of $A(f_0)$ (see I, §16), it is clear that there exists a sequence of quasi-linear functions $f_n(x, y)$ (see I, §15) such that $f_n - a \to f_0$ on $\mathfrak{I}$. Obviously each $f_n(x, y)$ is ACT on $\mathfrak{I}$ for $n = 1, 2, \cdots$, and hence satisfies III, (7.2). From III, §8 it follows that $f_0(x, y)$ also satisfies III, (7.2)—that is, any function which is BVT on $\mathfrak{I}$ satisfies III, (7.2).

10. Again, assume that $f_n - a \to f_0$ on $\mathfrak{I}$. From I, (17.7), we infer that

$$T_x(f_n - f_0) \leq T_x(f_n - \pi(f_n)x) + T_x(\pi(f_n)x - \pi(f_0)x) + T_x(f_0 - \pi(f_0)x).$$

Now (see III, (8.1)) it is clear that

$$\lim T_x(\pi(f_n)x - \pi(f_0)x) = \lim |\, \pi(f_n) - \pi(f_0)\,| \cdot |\mathfrak{I}| = 0.$$

Using III, (8.2), §9, and these inequalities, we conclude that

$$\limsup T_x(f_n - f_0) \leq 2^{3/2}\sigma(f_0)\, A(f_0)^{\frac{1}{2}}\{A(f_0) - \omega(f_0)\}^{\frac{1}{2}}. \tag{10.1}$$

11. As we observed in I, §18, the fact that $f_n - a \to f_0$ on $\mathfrak{I}$ implies that $f_n - a \to f_0$ on any subinterval Δ of $\mathfrak{I}$. Denote by $T_x(f, \Delta)$, $A(f, \Delta)$, $\sigma(f, \Delta)$, $\omega(f, \Delta)$ the interval functions defined with respect to the function $f(x, y)$ for $(x, y) \in \Delta$ as the corresponding functions $T_x(f)$, $A(f)$, $\sigma(f)$, $\omega(f)$ are defined in terms of the function $f(x, y)$ for $(x, y) \in \mathfrak{I}$ (see I, §§12-16, III, §6). Then if $f_n - a \to f_0$ on $\mathfrak{I}$, it follows from III, (10.1) that

$$\limsup T_x(f_n - f_0, \Delta) \leq 2^{3/2}\sigma(f_0, \Delta)\, A(f_0, \Delta)^{\frac{1}{2}}\{A(f_0, \Delta) - \omega(f_0, \Delta)\}^{\frac{1}{2}}. \tag{11.1}$$

Using the continuity of $f(x, y)$, the reader will verify that $T_x(f, \Delta)$ is an additive interval function.

12. Now let $D(\mathfrak{I})$ be any subdivision of $\mathfrak{I}$ (see II, §1). If M is any positive integer, denote by $D(M)$ that subsystem of intervals Δ contained in $\mathfrak{I}$ for which $\sigma(f_0, \Delta) \geq M$; denote by $\overline{D}(M)$ the system of intervals in $D(\mathfrak{I})$ and not in $D(M)$. We now assume that $f_0(x, y)$ is ACT on $\mathfrak{I}$; then it is clearly ACT on any subinterval of $\mathfrak{I}$ (see I, §13), and consequently (see I, §17)

$$A(f_0, \Delta) = \iint_{\Delta} [p(x, y)^2 + q(x, y)^2 + 1]^{\frac{1}{2}}\, dx\, dy. \tag{12.1}$$

From III, (6.1), (7.1) and the lemma in II, §10, it follows easily that

$$\omega(f_0, \Delta) \leq A(f_0, \Delta), \qquad U(\mathfrak{I}, \omega) = A(f_0). \tag{12.2}$$

From I, §17 one obtains

$$T_x(f_n - f_0, \Delta) \leq T_x(f_n, \Delta) + T_x(f_0, \Delta) \leq A(f_n, \Delta) + A(f_0, \Delta) \quad (\Delta \subset \mathfrak{I}).$$

Using relations (12.1), (12.2) we conclude that

$$\begin{aligned} M \cdot \sum_{\Delta \in D(M)} |\Delta| &\leq \sum_{\Delta \in D(M)} |\Delta| \cdot \sigma(f_0, \Delta) \\ &= \omega(f_0, D(M)) \leq A(f_0, D(M)) \leq A(f_0). \end{aligned} \tag{12.3}$$

Assume again that $f_n - a \to f_0$ on $\mathfrak{I}$; then $f_n - a \to f_0$ on Δ where Δ is any subinterval of $\mathfrak{I}$, so that (see (12.1))

$$\begin{aligned} \limsup T_x(f_n - f_0, D(M)) &\leq \limsup A(f_n, D(M)) + A(f_0, D(M)) \\ &\leq 2A(f_0, D(M)) = 2 \sum_{\Delta \in D(M)} \iint_{\Delta} [p^2 + q^2 + 1]^{\frac{1}{2}}\, dx\, dy. \end{aligned} \tag{12.4}$$

From relations (12.3) and (12.4) and the absolute continuity of the integral, it follows that, for any positive number ϵ, there exists a positive number M_ϵ, depending solely upon ϵ, such that for any subdivision $D(\mathfrak{Z})$ it is true that $\limsup T_x(f_n - f_0, D(M_\epsilon)) < \epsilon$.

13. Assume that $f_n - a \to f_0$ on $\mathfrak{Z}$ and that $f_0(x, y)$ is ACT on $\mathfrak{Z}$. Then if $D(\mathfrak{Z})$ be any subdivision, it follows from III, §§11, 12 and the lemma of Schwarz that

$$\begin{aligned}\limsup T_x(f_n - f_0) &= \limsup T_x(f_n - f_0, D(\mathfrak{Z})) \\ &\leq \limsup T_x(f_n - f_0, D(M_\epsilon)) + \limsup T_x(f_n - f_0, \overline{D}(M_\epsilon)) \\ &< \epsilon + 2^{3/2} M_\epsilon A(f_0)^{\frac{1}{2}} \{A(f_0) - \omega(f_0, D(\mathfrak{Z}))\}^{\frac{1}{2}}.\end{aligned}$$

In view of III, (12.2), the right member of this inequality is arbitrarily small if ϵ and $D(\mathfrak{Z})$ are properly chosen. Thus $\lim T_x(f_n - f_0) = 0$. Interchanging the rôles of x and y, one obtains $\lim T_y(f_n - f_0) = 0$. This proves the assertion made in III, §6.

14. We next make a proof for the theorem in I, §25. First, observe that, from the result of Radó quoted in I, §24, it follows that each of the Jacobians for the triple $\mathfrak{x}(u, v)$ is summable on $\mathfrak{Z}$. Also, conditions I, (25.4)-(25.7) imply that $\mathfrak{x}(u, v)$ is absolutely continuous AJ (see I, §21). From condition I, (25.6) and this fact, it follows that

$$A(\mathfrak{x}_n) = \iint_{\mathfrak{Z}} [J(x_n; u, v)^2 + J(y_n; u, v)^2 + J(z_n; u, v)^2]^{\frac{1}{2}}\, du\, dv \qquad (n = 1, 2, 3, \cdots);$$

$$A(\mathfrak{x}) = \iint_{\mathfrak{Z}} [J(x; u, v)^2 + J(y; u, v)^2 + J(z; u, v)^2]^{\frac{1}{2}}\, du\, dv. \tag{14.1}$$

We shall first prove this theorem, making the following additional assumption.

(14.2) Each of the functions $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$ $(n = 1, 2, \cdots)$ is quasi-linear on $\mathfrak{Z}$ (see I, §15).

If Δ be any interval contained in $\mathfrak{Z}$, define

$$\phi_n(\Delta) = \iint_{\Delta} |J(x_n; u, v)|\, du\, dv, \qquad \psi_n(\Delta) = \iint_{\Delta} |J(y_n; u, v)|\, du\, dv,$$

$$\chi_n(\Delta) = \iint_{\Delta} |J(z_n; u, v)|\, du\, dv; \qquad \phi_0(\Delta) = \iint_{\Delta} |J(x; u, v)|\, du\, dv,$$

$$\psi_0(\Delta) = \iint_{\Delta} |J(y; u, v)|\, du\, dv, \qquad \chi_0(\Delta) = \iint_{\Delta} |J(z; u, v)|\, du\, dv.$$

Since each of these interval functions is additive, it follows that (see II, §3)

$$(14.3)\quad U(\mathfrak{I}, \phi_n) = \phi_n(\mathfrak{I}), \qquad U(\mathfrak{I}, \psi_n) = \psi_n(\mathfrak{I}), \qquad U(\mathfrak{I}, \chi_n) = \chi_n(\mathfrak{I}) \qquad (n = 0, 1, 2, \cdots).$$

So it is obvious that these interval functions satisfy conditions II, (8.1), (8.3), (8.4). From the lemma in II, §10 and the relations (14.1), we obtain

$$U(\mathfrak{I}, \omega_n) = A(\mathfrak{x}_n) \qquad (n = 1, 2, 3, \cdots);$$

$$U(\mathfrak{I}, \omega_0) = A(\mathfrak{x}).$$

From assumption I, (25.7) it follows that condition II, (8.5) is fulfilled. In view of conditions I, (25.1)-(25.5) and the assumption (14.2), the hypotheses of a generalized lemma of McShane [9] are satisfied; consequently, for each interval δ contained in $\mathfrak{I}$, there exists a measurable set V_n contained in δ such that

$$\lim \iint_{V_n} J(x_n\,;\, u, v)\, du\, dv = \iint_{\delta} J(u, v)\, du\, dv.$$

Hence

$$\liminf \phi_n(\delta) \geq \Omega(\delta), \qquad \text{where } \Omega(\delta) = \left| \iint_{\delta} J(u, v)\, du\, dv \right| \qquad (\delta \subset \mathfrak{I}).$$

Now let Δ be any interval contained in $\mathfrak{I}$, and let $D(\Delta)$ be any subdivision; then clearly

$$\liminf \phi_n(\Delta) = \liminf \phi_n(D(\Delta)) \geq \Omega(D(\Delta)) \qquad (\Delta \subset \mathfrak{I}).$$

But from the lemma in II, §10, we find that

$$U(\Delta, \Omega) = \iint_{\Delta} |\, J(x;\, u, v)\,|\, du\, dv = \phi_0(\Delta) \qquad (\Delta \subset \mathfrak{I}).$$

Thus $\liminf \phi_n(\Delta) \geq \phi_0(\Delta)$ for $\Delta \subset \mathfrak{I}$. Similar inequalities follow for the interval functions $\psi_n(\Delta)$ and $\chi_n(\Delta)$. Therefore, condition II, (8.2) is satisfied. From the corollary in II, §8, the assertions I, (25.8) now follow, in view of relations (14.3).

15. We now prove the theorem in I, §25, dropping the assumption III, (14.2). From the definition of the area (see I, §§22, 23), it follows that there exists, for each n, a sequence of triples of quasi-linear functions

$$(15.1)\quad \mathfrak{x}_{mn}(u, v) = (x_{mn}(u, v), y_{mn}(u, v), z_{mn}(u, v)) \quad ((u, v) \in \mathfrak{I};\ m = 1, 2, 3, \cdots),$$

such that the $x_{mn}(u, v)$, $y_{mn}(u, v)$, $z_{mn}(u, v)$ converge uniformly on $\mathfrak{I}$ to $x_n(u, v)$, $y_n(u, v)$, $z_n(u, v)$, respectively, and $A(\mathfrak{x}_{mn})$ converges to $A(\mathfrak{x}_n)$ as m tends to ∞.

Since by I, (25.6) each of the $\mathfrak{x}_n(u, v)$ is absolutely continuous AJ, it follows that, for each n, the triples $\mathfrak{x}_{mn}(u, v)$, $\mathfrak{x}_n(u, v)$ $(m = 1, 2, 3, \cdots)$ satisfy the assumptions of the theorem in I, §25 and the additional assumption III, (14.2) made in the preceding section. Hence, from the preceding section, we conclude that

$$\lim \iint_{\mathfrak{J}} | J(x_{mn} ; u, v) | \, du \, dv = \iint_{\mathfrak{J}} | J(x_n ; u, v) | \, du \, dv,$$

$$\lim \iint_{\mathfrak{J}} | J(y_{mn} ; u, v) | \, du \, dv = \iint_{\mathfrak{J}} | J(y_n ; u, v) | \, du \, dv, \tag{15.2}$$

$$\lim \iint_{\mathfrak{J}} | J(z_{mn} ; u, v) | \, du \, dv = \iint_{\mathfrak{J}} | J(z_n ; u, v) | \, du \, dv,$$

$$(n = 1, 2, 3, \cdots).$$

Thus, for each n, there exists a triple $\mathfrak{x}_n^*(u, v)$ in the set (15.1) for which the following inequalities hold simultaneously.

$$\left.\begin{aligned} &|| \, \mathfrak{x}_n^*(u, v) - \mathfrak{x}_n(u, v) \, || < n^{-1} \qquad ((u, v) \epsilon \mathfrak{J}, \ n = 1, 2, 3, \cdots); \\ &\left| \iint_{\mathfrak{J}} | J(x_n^* ; u, v) | \, du \, dv - \iint_{\mathfrak{J}} | J(x_n ; u, v) | \, du \, dv \right| \\ &\left| \iint_{\mathfrak{J}} | J(y_n^* ; u, v) | \, du \, dv - \iint_{\mathfrak{J}} | J(y_n ; u, v) | \, du \, dv \right| \\ &\left| \iint_{\mathfrak{J}} | J(z_n^* ; u, v) | \, du \, dv - \iint_{\mathfrak{J}} | J(z_n ; u, v) | \, du \, dv \right| \end{aligned}\right\} < n^{-1} \tag{15.3}$$

$$(n = 1, 2, 3, \cdots).$$

Now clearly the triples $\mathfrak{x}_n^*(u, v)$, $\mathfrak{x}(u, v)$ satisfy the assumptions of the theorem in I, §25 and the additional assumption III, (14.2), so that

$$\lim \iint_{\mathfrak{J}} | J(x_n^* ; u, v) | \, du \, dv = \iint_{\mathfrak{J}} | J(x; u, v) | \, du \, dv,$$

$$\lim \iint_{\mathfrak{J}} | J(y_n^* ; u, v) | \, du \, dv = \iint_{\mathfrak{J}} | J(y; u, v) | \, du \, dv, \tag{15.4}$$

$$\lim \iint_{\mathfrak{J}} | J(z_n^* ; u, v) | \, du \, dv = \iint_{\mathfrak{J}} | J(z; u, v) | \, du \, dv.$$

Relations (15.3) and (15.4) imply the conclusion I, (25.8).

16. To verify the corollary in I, §26, we notice that there always exists a sequence of triples $\mathfrak{x}_n(u, v)$ of quasi-linear functions which satisfy the hypotheses of the theorem in I, §25, provided the triple $\mathfrak{x}(u, v)$ is absolutely continuous AJ. For these $\mathfrak{x}_n(u, v)$, we have, using the notation of the corollary,

$$A(x\text{-}\mathfrak{x}_n) = \iint_{\mathfrak{I}} | J(x_n ; u, v) | \, du \, dv,$$

$$A(y\text{-}\mathfrak{x}_n) = \iint_{\mathfrak{I}} | J(y_n ; u, v) | \, du \, dv,$$

$$A(z\text{-}\mathfrak{x}_n) = \iint_{\mathfrak{I}} | J(z_n ; u, v) | \, du \, dv \qquad (n = 1, 2, 3, \cdots).$$

From the conclusion I, (25.8) and the facts stated in I, §24, the corollary in I, §26 is now immediate.

17. As we noted above, there always exists a sequence of triples $\mathfrak{x}_n(u, v)$ of quasi-linear functions which satisfy the hypotheses of the theorem in I, §25, provided $\mathfrak{x}(u, v)$ is absolutely continuous AJ. Now the transformations defined by the pairs

$$(17.1) \qquad \begin{cases} y = y_n(u, v), \\ z = z_n(u, v), \end{cases} \quad \begin{cases} z = z_n(u, v), \\ x = x_n(u, v), \end{cases} \quad \begin{cases} x = x_n(u, v), \\ y = y_n(u, v), \end{cases} \qquad (u, v) \in \mathfrak{I},$$

are each in the class K_3, since a quasi-linear function is Lipschitzian (see I, §28). From the remark at the end of II, §8, it follows that the relations I, (25.8) also hold on every subinterval of $\mathfrak{I}$. But now, in view of I, §25, it is clear that each of the sets of transformations in (17.1) taken together with the corresponding transformation in I, (27.1) satisfies the assumptions of the closure theorem for the class K_3, stated in I, §28; from this theorem follows the corollary in I, §27.

Bibliography

1. C. R. Adams and J. A. Clarkson, *On convergence in variation*, Bulletin of the American Mathematical Society, vol. 40(1934), pp. 413–417.
2. C. R. Adams and Hans Lewy, *On convergence in length*, this Journal, vol. 1(1935), pp. 19–26.
3. G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities*, Cambridge, 1934.
4. E. J. McShane, *On a certain inequality of Steiner*, Annals of Mathematics, second series, vol. 33(1932), pp. 125–138.
5. Tibor Radó, *Sur l'aire des surfaces courbes*, Acta Litterarum ac Scientiarum, vol. 3(1927), pp. 131–169.
6. Tibor Radó, *Sur le calcul de l'aire des surfaces courbes*, Fundamenta Mathematicae, vol. 10(1927), pp. 197–210.
7. Tibor Radó, *On the derivative of the Lebesgue area of continuous surfaces*, Fundamenta Mathematicae, vol. 30(1938), pp. 34–39.

8. TIBOR RADÓ, *Über das Flächenmass rektifizierbarer Flächen*, Mathematische Annalen, vol. 100(1928), pp. 445–479.
9. TIBOR RADÓ, *On a lemma of McShane*, Annals of Mathematics, vol. 42(1941), pp. 73–83.
10. TIBOR RADÓ, *On the problem of Plateau*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Berlin, 1933.
11. T. RADÓ AND P. REICHELDERFER, *Note on an inequality of Steiner*, Bulletin of the American Mathematical Society, vol. 47(1941), pp. 102–108.
12. T. RADÓ AND P. REICHELDERFER, *A theory of absolutely continuous transformations in the plane*, Transactions of the American Mathematical Society, vol. 49(1941), pp. 258–307.
13. STANISLAW SAKS, *Theory of the Integral*, Warszawa-Lwów, 1937.

THE OHIO STATE UNIVERSITY AND THE UNIVERSITY OF CHICAGO.

THE ABSOLUTE CESÀRO SUMMABILITY OF TRIGONOMETRICAL SERIES

By Fu Traing Wang

1. The absolute Cesàro summability of series has been defined by Kogbetliantz [9] and treated by many other authors; it is related to Cesàro summability as absolute convergence is related to ordinary convergence. Bosanquet [1], [2], [3] applied this kind of summability to Fourier series and proved that the property of bounded variation of the Cesàro mean of the function is the necessary and sufficient condition that its Fourier series be absolutely Cesàro summable at a point. Hyslop [6], [7], Cooper [5], Chow [4], and Randels [12] discussed the analogous problem for derived Fourier series, conjugate series, Fourier integrals and power series.

Put

$$A_n = A_n(x) = A_n \cos nx + b_n \sin nx,$$

and

$$\sigma_n^\alpha = \frac{1}{\binom{n+\alpha}{n}} \sum_{\nu=0}^{n} \binom{n-\nu+\alpha}{n-\nu} A_\nu .$$

A series $\sum A_n$ is said to be absolutely summable (C, α) or, simply, summable $|C, \alpha|$ provided the series $\sum (\sigma_n^\alpha - \sigma_{n-1}^\alpha)$ is absolutely convergent. The object of this paper is to prove the following theorem and find its best possible condition.

Theorem. *If the series*

$$\sum_{n=2}^{\infty} (a_n^2 + b_n^2)(\log n)^{1+\epsilon} \qquad (\epsilon > 0) \tag{1.1}$$

is convergent, then the trigonometrical series

$$\sum_{n=0}^{\infty} (a_n \cos nx + b_n \sin nx) \tag{1.2}$$

is summable $|C, \alpha|$ $(\alpha > \frac{1}{2})$ *almost everywhere.*

First, we can prove that the trigonometrical series

$$\sum_{n=2}^{\infty} (n \log n)^{-1} \cos 2^n x$$

when $\sum_{n=2}^{\infty} a_n^2 \log n$ converges is non-absolutely Cesàro summable at a set of points

Received March 26, 1942.

of positive measure and, second, we can construct a trigonometrical series which satisfies the condition of the theorem and is non-summable $| C, \alpha |$ $(0 < \alpha < \frac{1}{2})$ almost everywhere.

2. We use A or O for a constant which depends only on α and is different in different occurrences. Before proving the theorem, we require a number of lemmas.

LEMMA 1 (Kogbetliantz [9]). *If*

$$\zeta_n^\alpha = \zeta_n^\alpha(x) = \frac{1}{\binom{\alpha+n}{n}} \sum_{\nu=1}^{n} \binom{n-\nu+\alpha-1}{n-\nu} \nu A_\nu ,$$

then

$$n(\sigma_n^\alpha - \sigma_{n-1}^\alpha) = \zeta_n^\alpha .$$

LEMMA 2 (Kogbetliantz [9]). *If a series* $\sum A_n$ *is summable* $| C, \alpha |$, *then it is summable* $| C, \beta |$ *for* $\beta > \alpha$.

LEMMA 3. *If* $\frac{1}{2} < \alpha < 1$, *then*

$$\sum_{n=\nu}^{\infty} \binom{n-\nu+\alpha-1}{n-\nu}^2 \frac{(\log n)^{1+\epsilon}}{n^{1+2\alpha}} = O\left(\frac{(\log \nu)^{1+\epsilon}}{\nu^2}\right).$$

This is easily seen from Stirling's formula.

LEMMA 4 (Rajchman and Saks [11], Fubini's theorem). *If* $\{\phi_n(t)\}$ *is a sequence of positive monotone increasing functions defined in the interval* (a, b) *and* $\sum \phi_n(b)$ *converges, then* $\sum \phi_n'(t)$ *is convergent almost everywhere.*

Proof of the theorem. By Lemma 2 we may suppose that $\frac{1}{2} < \alpha < 1$ and by Lemma 1 and Lemma 4 it suffices to show that the series

$$\sum_{n=1}^{\infty} \int_0^{2\pi} \left| \frac{\zeta_n^\alpha(x)}{n} \right| dx \tag{2.1}$$

converges. Now we have

$$\int_0^{2\pi} | \zeta_n^\alpha(x) | \, dx \le \frac{A}{n^\alpha} \left\{ \sum_{\nu=1}^{n} \binom{n-\nu+\alpha-1}{n-\nu}^2 \nu^2 (a_\nu^2 + b_\nu^2) \right\}^{\frac{1}{2}}. \tag{2.2}$$

By Lemma 3 and (1.1) we have

$$\sum_{n=2}^{m} \frac{1}{n^{\alpha+1}} \left\{ \sum_{\nu=1}^{n} \binom{n-\nu+\alpha-1}{n-\nu}^2 \nu^2 (a_\nu^2 + b_\nu^2) \right\}^{\frac{1}{2}}$$

$$\leq \left\{ \sum_{n=2}^{m} \frac{1}{n(\log n)^{1+\epsilon}} \right\}^{\frac{1}{2}} \left\{ \sum_{n=2}^{m} \frac{(\log n)^{1+\epsilon}}{n^{1+2\alpha}} \sum_{\nu=1}^{n} \binom{n-\nu+\alpha-1}{n-\nu}^2 \nu^2 (a_\nu^2 + b_\nu^2) \right\}^{\frac{1}{2}} \tag{2.3}$$

$$\leq A \left\{ \sum_{\nu=1}^{m} \nu^2 (a_\nu^2 + b_\nu^2) \sum_{n=\nu}^{m} \binom{n-\nu+\alpha-1}{n-\nu}^2 \frac{(\log n)^{1+\epsilon}}{n^{1+2\alpha}} \right\}^{\frac{1}{2}}$$

$$\leq A \left\{ \sum_{\nu=1}^{m} (a_\nu^2 + b_\nu^2)(\log n)^{1+\epsilon} \right\}^{\frac{1}{2}} \leq A.$$

By (2.2) and (2.3) we have

$$\sum_{n=2}^{m} \int_0^{2\pi} \left| \frac{\zeta_n^{\alpha}(x)}{n} \right| dx \leq A.$$

This proves that (2.1) is convergent.

3. Lemma 5. *If $\{F_m(x)\}$ is a sequence of integrable functions defined in the interval $(0, 2\pi)$ and satisfies the conditions*

$$0 \leq F_m(x) \leq A_1 \log_3 m \qquad \text{for all } x \text{ and } m \tag{3.1}$$

and

$$\int_0^{2\pi} F_m(x)dx \geq A_2 \log_3 m \qquad \text{for } m \geq m_1, \tag{3.2}$$

then there exists a set of points E of positive measure such that $F_m(x) > \log_4 m$ for $x \in E$ and infinitely many values of m.

We use the following notation: $\log_r m = \log(\log_{r-1} m)$, $\prod C_\nu$ is the set of points which the C's have in common and $|C|$ is the measure of the set C.

If Lemma 5 is not true, then $F_m(x) \leq \log_4 m$ for $m \geq N(x)$ and x belongs to a set of points I of measure 2π. Let C_m be the set of points of x in I such that $F_m(x) \leq \log_4 m$ and $C_m = \prod_{\nu=m}^{\infty} C_\nu$; then $C_m \subset C_{m+1}$. If C is the outer limiting set of the sequence $\{C_m\}$, then $I = C$. Given a positive number δ we can find a positive integer k such that $|C_k| > 2\pi - \delta$. Hence $F_m(x) \leq \log_4 m$ for $x \in C_k$ and $m \geq k$; then

$$\int_{C_k} F_m(x)dx \leq 2\pi \log_4 m,$$

for $n \geq k$ and

$$\int_0^{2\pi} F_m(x)dx \leq 2\pi \log_4 m + \int_{C(C_k)} F_m(x)dx.$$

By (3.1) and (3.2) we have

$$A_2 \log_3 m = 2\pi \log_4 m + A_1 \delta \log_3 m$$

for $m \geq \max(k, m_1)$; this inequality is seen to be impossible by taking m sufficiently large and then δ small. Thus the lemma is proved.

In order to prove that the series $\sum_{n=2}^{\infty} (n \log n)^{-1} \cos 2^n x$ is non-summable $|C, \alpha|$ at a set of points of positive measure, we may suppose that $\alpha \geq 1$ and put

$$r_n = \left[\frac{\log n}{\log 2}\right], \qquad f_n(x) = \left| \sum_{\nu=2}^{r_n} \binom{n - 2^\nu + \alpha - 1}{n - 2^\nu} \frac{2^\nu}{\nu \log \nu} \cos 2^\nu x \right|.$$

Let e_ν be the set of points in $(0, 2\pi)$ for which $\cos 2^\nu x$ is positive and $E_n = \prod_{\nu=r_{n-l}}^{r_n} e_\nu$; then $|E_n| = \pi 2^{-l}$ and

$$\int_{E_n} \cos 2^{r_n - 1} x dx = 2^{3/2 - l} \tag{3.3}$$

Now

$$\begin{aligned}\int_0^{2\pi} f_n(x)dx &\geq \int_{E_n} f_n(x)dx \\ &\geq \int_{E_n} \left| \sum_{\nu=r_{n-l}}^{r_n} \binom{n - 2^\nu + \alpha - 1}{n - 2^\nu} \frac{2^\nu}{\nu \log \nu} \cos 2^\nu x \right| dx \\ &\quad - \int_{E_n} \left| \sum_{\nu=2}^{r_{n-l-1}} \binom{n - 2^\nu + \alpha - 1}{n - 2^\nu} \frac{2^\nu}{\nu \log \nu} \cos 2^\nu x \right| dx.\end{aligned}$$

By Stirling's formula and

$$A_1(n - 2^\nu)^{\alpha-1} \leq \binom{n - 2^\nu + \alpha - 1}{n - 2^\nu} \leq A_2(n - 2^\nu)^{\alpha-1},$$

we have

$$\begin{aligned}\int_0^{2\pi} f_n(x)dx \geq A_1(n - 2^{r_n-1})^{\alpha-1} \frac{2^{r_n-1}}{\log n \log_2 n} \int_{E_n} \cos 2^{r_n-1} x dx \\ - A_2 |E_n| n^{\alpha-1} \sum_{\nu=2}^{r_{n-l-1}} \frac{2^\nu}{\nu \log \nu}.\end{aligned}$$

If we take $r_n > 2l$, then

$$\begin{aligned}\sum_{\nu=2}^{r_{n-l-1}} \frac{2^\nu}{\nu \log \nu} &\leq \int_2^{r_n-l} \frac{2^u}{u \log u} du \\ &= O\left(\frac{2^{r_n-l}}{(r_n - l) \log (r_n - l)}\right) = O\left(\frac{n}{\log n \log_2 n} 2^{-l}\right).\end{aligned}$$

Hence, for $n \geq m_1$,

$$(3.4)\quad \int_0^{2\pi} f_n(x)dx \geq 2^{-l} \frac{n^\alpha}{\log n \log_2 n} (A_1 - A_2 2^{-l}) \geq A \frac{n^\alpha}{\log n \log_2 n},$$

if we take l sufficiently large and fixed.

Put

$$F_m(x) = \sum_{n=2}^{m} \frac{f_n(x)}{n\binom{n+\alpha}{n}},$$

then by (3.4)

$$\int_0^{2\pi} F_m(x)dx \geq A \sum_{n=m_1}^{m} \frac{1}{n^{1+\alpha}} \int_0^{2\pi} f_n(x)dx$$

$$\geq A \sum_{n=m_1}^{m} \frac{1}{n \log n \log_2 n} \geq A \log_3 m, \qquad \text{for } m \geq m_1.$$

It is easily seen that $0 \leq F_m(x) \leq A_2 \log_3 m$ for all x and m. Hence, by Lemma 5, a set of points E of positive measure exists for which the series

$$\sum_{n=2}^{\infty} \frac{1}{n\binom{n+\alpha}{n}} \left| \sum_{\nu=2}^{r_n} \binom{n-2^\nu+\alpha-1}{n-2^\nu} \frac{2^\nu}{\nu \log \nu} \cos 2^\nu x \right|$$

diverges at all points of it. That is to say, the series

$$\sum_{n=2}^{\infty} \frac{1}{n \log n} \cos 2^n x$$

is non-summable $| C, \alpha |$ at all points of E.

4. On the other hand, we can prove that the series

$$(4.1)\quad \sum_{n=10}^{\infty} \frac{\cos nx}{n^{1-\alpha} \log n \log_2 n} \qquad (0 < \alpha < \tfrac{1}{2})$$

which satisfies the condition of the theorem is non-summable $| C, \alpha |$ almost everywhere by the following

LEMMA 6 (Kogbetliantz [9]). *If the series* $\sum A_n$ *is summable* $| C, \alpha |$, *then* $\sum n^{-\alpha} A_n$ *is absolutely convergent.*

In fact, if the series (4.1) is summable $| C, \alpha |$ at a set of points of positive measure, then by Lusin's theorem [8] and Lemma 6 the series

$$\sum \frac{\cos nx}{n \log n \log_2 n}$$

converges everywhere. This is impossible for $x = 0$.

In conclusion, I state, without proof, the following results.

(I) If $\sum (a_n^2 + b_n^2) (\log n)^{2+\epsilon}$ is convergent, then the series (1.2) is summable $| C, \frac{1}{2} |$ almost everywhere.

(II) If $0 < \alpha < \frac{1}{2}$ and $\sum (a_n^2 + b_n^2) n^{1-2\alpha} (\log n)^{1+\epsilon}$ converges, then the series (1.2) is summable $| C, \alpha |$ almost everywhere. The series (4.1) shows that the positive number ϵ cannot be omitted in this case.

Bibliography

1. L. S. Bosanquet, *The absolute Cesàro summability of a Fourier series*, Proceedings of the London Mathematical Society, vol. 41(1936), pp. 517–528.
2. L. S. Bosanquet, *Note on the absolute summability (C) of a Fourier series*, Journal of the London Mathematical Society, vol. 11(1936), pp. 11–15.
3. L. S. Bosanquet and J. M. Hyslop, *On the absolute summability of the allied series of a Fourier series*, Mathematische Zeitschrift, vol. 42(1937), pp. 489–512.
4. H. C. Chow, *On the absolute summability (C) of power series*, Journal of the London Mathematical Society, vol. 14(1939), pp. 101–112.
5. J. L. B. Cooper, *The absolute Cesàro summability of Fourier integrals*, Proceedings of the London Mathematical Society, vol. 45(1939), pp. 425–439.
6. J. M. Hyslop, *On the absolute summability of Fourier series*, Proceedings of the London Mathematical Society, vol. 43(1937), pp. 475–483.
7. J. M. Hyslop, *On the absolute summability of the successively derived series of a Fourier series and its allied series*, Proceedings of the London Mathematical Society, vol. 46(1939), pp. 55–80.
8. E. W. Hobson, *The Theory of Functions of a Real Variable and the Theory of Fourier's Series*, vol. 2, 1926, p. 549.
9. E. Kogbetliantz, *Sur les séries absolument sommables par la méthode des moyennes arithmétiques*, Bulletin des Sciences Mathématiques, (2), vol. 49(1925), pp. 234–256.
10. E. Kogbetliantz, *Sommation des séries et intégrales divergentes par les moyennes arithmétiques et typiques*, Mémorial des Sciences Mathématiques, no. 51, 1931.
11. A. Rajchman and S. Saks, *Sur la dérivabilité des fonctions monotones*, Fundamenta Mathematicae, vol. 4(1923), pp. 204–213.
12. W. C. Randels, *On the absolute summability of Fourier series* III, this Journal, vol. 7(1940), pp. 204–207.

National University of Chekiang,
Kweichow, China.

THEORY OF EQUIVALENCE RELATIONS

By Oystein Ore

The present paper contains an extensive analysis of the theory of equivalence relations. A large number of new results have been obtained and it seems of particular interest that several of these connect rather diverse mathematical fields. But this theory appears to be of importance also for various other reasons. It may be mentioned, for instance, that a greater part of the theory consists in a study of the structure of all equivalence relations over a set. This structure does not satisfy the Dedekind condition and it is of a sufficiently general type to afford a useful example for arbitrary structure theory. Several problems solved for the structure of equivalence relations give indications of suitable methods of attack for analogous problems in the general case. As another justification for the theory of equivalence relations one may mention the fact that it yields an example for the much more general theory of mathematical relations which I have been studying for some time.

The main contents of the paper are as follows. In the first chapter one finds the simplest properties of *equivalence relations*, their representation as *partitions* or *fields of sets* over the basic set S, and it is shown that they form a complete structure. All these results were obtained by Garrett Birkhoff for the case of a finite set S, but the extension to arbitrary sets entails no particular difficulties. D. König made the important observation that there is associated a special so-called *pair graph* with the union of any two equivalence relations. Next the *Dedekind relation* and the *distributive law* are studied in the structure of equivalence relations and it is of interest that the results can be interpreted as properties of the union graph of two relations; for instance, the Dedekind relation is satisfied in one form if and only if the graph is a *tree*. The law of isomorphism is discussed by means of the Dedekind law. This leads to problems investigated by the Dubreils on the theorem of Jordan-Hölder and isomorphic chain refinements.

The structure of equivalence relations is *complemented* or even *completely complemented*. Furthermore, it possesses a special type of complements which I have called *Dedekind complements*. The construction of a Dedekind complement corresponds to a choice of representatives in the sets of the partition. The well-known problem of finding *common representatives* for two partitions may therefore be formulated structurally as a determination of common Dedekind complements. A few illustrations of the application of the theory to groups are given. It is shown that the existence of common representatives for right and left co-set expansions depends on the fact that the right and left co-set expansions always define so-called *commuting equivalence relations*.

In the third chapter some connections between partitions and *correspondences*

Received April 6, 1942.

are derived. These investigations have applications also to other problems not mentioned here. The *automorphisms* and *endomorphisms* of an equivalence relation are determined and finally some properties of *permutation representations* of structures are indicated.

In the last chapter the structure of equivalence relations is characterized as a *geometric system* with points and lines with rather peculiar properties. It is also shown that the *group of automorphisms* of the structure of all equivalence relations is the symmetric group on the basic set S, a result also due to Garrett Birkhoff in the finite case. More difficult to establish is the fact that the structure of equivalence relations has no endomorphisms. This is shown to be a consequence of a general property of relatively complemented structures.

Chapter 1

1. **Equivalence relations and partitions.** In the following we shall denote by S some set which shall remain fixed. An *equivalence relation* E in S is a *binary relation*

$$a\ E\ b$$

between two elements a and b in S, defined by the four properties:

(1) *Determination.* For any pair of elements a and b the relation $a\ E\ b$ either holds or does not hold.

(2) *Reflexivity.* For any a one has $a\ E\ a$.

(3) *Symmetry.* When $a\ E\ b$, then $b\ E\ a$.

(4) *Transitivity.* When $a\ E\ b$ and $b\ E\ c$, then $a\ E\ c$.

The special equivalence relation

$$a\ U\ b$$

which is defined to hold for *any* pair of elements a and b will be called *the universal relation*. Similarly the equivalence relation

$$a\ I\ b$$

which holds only when $a = b$ will be called the *identity* or *unit relation*.

A *partition* P of the set S is a decomposition of S into subsets C_α such that every element in S belongs to one and only one set C_α. We shall call the sets C_α the *blocks* of the partition P and write $P = P(C_\alpha)$.

The connection between partitions and equivalence relations is expressed in the following well-known result.

THEOREM 1. *Any partition $P(C)$ defines an equivalence relation E in the set S when one puts $a\ E\ b$ whenever a and b belong to the same block C. Conversely, any equivalence relation E defines a partition $P(C_\alpha)$ where the blocks C_α consist of all elements equivalent to any given element a.*

Proof. The correctness of the first part of the theorem is seen immediately, and to prove the converse one need only observe that every element in S belongs to some set C_α and that two such sets C_α must be either identical or disjoint according to the axioms for an equivalence relation.

Among the special partitions of S we note the *universal partition* P_U in which S is the only block and the *complete* or *identical partition* P_I in which every block is a single element.

When the set S is a finite set with n elements, the number p_n of partitions or equivalence relations over S is also finite. This number p_n occurs in various other mathematical problems and it has been studied extensively by A. C. Aitken, E. T. Bell, L. F. Epstein and many other writers. For a more complete set of references the reader may be referred to a paper by Epstein [5], whose appendix contains an extensive bibliography on the numbers p_n and related numbers. The numbers p_n may be determined successively through the recursion formula

$$p_{n+1} = \sum_{i=0}^{n} \binom{n}{i} p_i .$$

The number $n!p_n$ has the characteristic function

$$e^{e^x-1}.$$

2. **Partitions defined by families of sets.** In the following we shall use the ordinary terminology of the theory of sets. The *sum*, *intersection* and *difference* of two subsets A and B of S will be denoted respectively by $A + B$, $A \cdot B$ and

$$A - B = A - A \cdot B.$$

A *family of subsets* will be denoted by $\phi(A_m)$ where the marks m may run through a set M of arbitrary cardinal number. For the sum and intersection of the sets in the family we shall write respectively

$$\Sigma_\phi = \sum_{m \subset M} A_m \; ; \qquad \Pi_\phi = \prod_{m \subset M} A_m .$$

The family of sets $\phi(A_m)$ will be said to *cover* S when every element in S belongs to at least one set A_m.

Every family of sets defines a certain partition of the basic set S as we shall now show. Let us denote the given family of subsets by

$$\Phi = \Phi(A_m) \qquad (m \in M), \tag{1}$$

where the index set M may be arbitrary. Now let a be any element of S. Among the sets A_m in the family (1) there will be certain ones A_{m_1} which contain a and certain others A_{m_2} which do not contain this element. The marks m_1 and m_2 will form two sets M_1 and M_2 which are complementary in M. Since a belongs to all A_{m_1} it must also belong to their intersection Π_{M_1}. Similarly, since a does

not belong to any of the sets A_{m_2}, it does not belong to their sum Σ_{M_2}. This shows that a is an element of the set

$$C_{M_1} = \Pi_{M_1} - \Pi_{M_1} \cdot \Sigma_{M_2} . \tag{2}$$

We shall agree that if M_1 is void then

$$C_0 = S - \Sigma_M$$

and similarly if M_2 is void

$$C_M = \Pi_M .$$

All elements in S corresponding to the same set of marks M_1 and M_2 must belong to the same set C_{M_1}. Furthermore, no element b with different sets of marks M_1' and M_2' can belong to this same C_{M_1}. If, namely, b belonged to the intersection Π_{M_1}, one would have $M_1' \supseteq M_1$. But if $M_1' \supset M_1$ then $M_2' \subset M_2$ and b is contained in the sum Σ_{M_2}. But in this case b is eliminated from C_{M_1} by the subtracted term in (2). This shows that the sets C_{M_1} form the blocks in a partition of the set S so that we have proved

THEOREM 2. *Let*

$$\Phi(A_m) \qquad (m \in M)$$

be a family of subsets of a set S where the marks m run through some set M of arbitrary cardinal number. Any such family of subsets defines a partition $P(C_{M_1})$ of S where the blocks C_{M_1} are defined by

$$C_{M_1} = \prod_{m_1 \in M_1} A_{m_1} - \Big(\prod_{m_1 \in M_1} A_{m_1}\Big)\Big(\sum_{m_2 \in M - M_1} A_{m_2}\Big) \tag{3}$$

and M_1 runs through all subsets of M.

The proof of Theorem 1 implies further that an equality

$$C_{M_1} = C_{M_1'}$$

can only hold if the two sets are void or if $M_1 = M_1'$. The expression (3) for the blocks C_{M_1} can also be written in the forms

$$C_{M_1} = \Pi_{M_1} - \sum_{m_2 \in M - M_1} (A_{m_2} \cdot \Pi_{M_1}) = \Pi_{M_1} (S - \Sigma_{M_2}).$$

Theorem 2 shows that every family of sets (1) generates a partition of S. Let us turn to the converse problem where a partition $P(C)$ of the set S is given. We wish to determine all families $\phi(A)$ which generate this partition of S in the manner indicated by Theorem 2. Let the number of blocks in the partition $P(C)$ be indicated by the cardinal number γ_0. This cardinal number can be raised arbitrarily to any γ greater than γ_0 by adjoining the void set 0 a sufficient number of times to the family $P(C)$. According to Theorem 2 the sets C must be associated in a one-to-one manner with the subsets of some set M. Therefore, let M be chosen as a set with cardinal number μ so that the number of subsets of M is 2^μ. If $2^\mu > \gamma_0$ one can adjust γ as indicated so that $2^\mu = \gamma$. A one-to-

one correspondence α between the sets in $P(C)$ and the subsets of M may then be established. After this correspondence has been defined one can write any C as a C_{M_1} where M_1 is a uniquely determined subset of M. When these indices M_1 have been introduced, the family $\phi(A_m)$ which generates the corresponding partition is determined since A must be the sum of those sets C_{M_1} for which M_1 contains the element m. Conversely, when the A_m are determined this way their corresponding partition is seen to be $P(C)$.

This construction shows that other families $\phi(A_m)$ generating $P(C)$ may be obtained through a different correspondence α and through sets M with different cardinal numbers. When the number γ_0 of blocks is finite, the smallest number of sets A_m in a family $\phi(A_m)$ generating the partition $P(C)$ is μ, where μ is the smallest integer such that $2^\mu \geq \gamma$.

3. **Complete fields of sets.** A family of sets $\phi(A_m)$ will be called a *field of sets* when it contains the sum, intersection and difference of any two of its sets. The field of sets is said to be *complete* when it contains the sum and intersection of any subfamily of its sets. Finally, the field of sets shall be called a *field of sets over* S when it contains the complement $\overline{A} = S - A$ of any of its sets. This may also be stated that the field will contain S. Any family of sets ϕ generates a complete field of sets obtained by adjoining to ϕ all sets derived from the given ones by the successive application of the operations of forming differences and sum and intersection of arbitrary subfamilies. If ϕ covers S the resulting field is a complete field of sets over S.

If F_1 and F_2 are two complete fields of sets over S, their common sets form the largest complete field of sets over S contained in both, the cross-cut $F_1 \cap F_2$ of F_1 and F_2. Similarly, the union $F_1 \cup F_2$ of F_1 and F_2 is the least complete field of sets over S containing both F_1 and F_2 and it is the complete field generated by the sets in F_1 and F_2. These concepts of union and cross-cut can be extended to an arbitrary finite or infinite number of fields so that we see that all complete fields over S form a complete structure. The all-element of this structure is the field of all subsets of S. The zero-element is the field consisting simply of S and the void set 0.

When a family of sets (1) is given, it is obvious that the sets C_{M_1} in (3) belong to the complete field of sets over S generated by the family. On the other hand, it is clear also that a family of sets consisting of all sums $\sum C$ of blocks in a partition form a complete field. This gives

THEOREM 3. *The complete field of sets over S generated by a family of sets ϕ in* (1) *consists of all sums $\sum C_{M_1}$, where the sets C_{M_1}, are the blocks* (3) *in the partition of S defined by Φ.*

As a corollary, there results

THEOREM 4. *A complete field of sets over S consists of all sums of blocks in a partition of S.*

This result also shows that each partition corresponds to a single complete field of sets over S and conversely. We shall return to this one-to-one correspondence a little later. In this connection it may be remarked that the previous construction of all families generating a given partition also determines all families generating a given complete field over S. A more detailed discussion of the methods for constructing the field of sets generated by a family of sets can be found in [6] and [16].

4. **The structure of equivalence relations.** The system of all partitions of a set S forms a partially ordered set when one writes $P_1 \supset P_2$ whenever the blocks in P_2 are obtained by subdivisions of the blocks in P_1. The zero element 0 of this partially ordered set, i.e., the partition contained in all others, is the complete partition of S and the all-element containing all others is the universal partition consisting only of the block S. In terms of equivalence relations one writes $E_1 \supset E_2$ when $a\ E_2\ b$ always implies $a\ E_1\ b$. We now prove Theorem 5. For a finite set S this theorem as well as the following Theorem 6 is given by G. Birkhoff [1; Theorems 18 and 21].

THEOREM 5. *The system of all partitions or all equivalence relations of a set S forms a complete structure.*

Proof. For any two partitions $P_1(C_1)$ and $P_2(C_2)$ the cross-cut $P_1 \cap P_2$ is the partition whose blocks are the intersections $C_1 \cdot C_2$ of the blocks in P_1 and P_2. For an arbitrary set of partitions $\{P_i\}$ there also exists a cross-cut $\bigwedge_i P_i$ whose blocks are all intersections $\prod_i C_i$ with one C_i from each partition P_i. One can also consider $\bigwedge_i P_i$ as the partition obtained by the construction of Theorem 2 applied to the family of all sets C_i from the various partitions. For equivalence relations E_1 and E_2 the cross-cut $E_1 \cap E_2$ is the equivalence relation

$$a\ E_1 \cap E_2\ b$$

which holds if and only if simultaneously

$$a\ E_1\ b, \qquad a\ E_2\ b.$$

The union of two or more partitions may also be characterized in several ways. The blocks $C_{1,2}$ in the partition $P_1 \cup P_2$ must contain the blocks of P_1 and P_2 as refinements. This shows that

$$C_{1,2} = \sum C_1 = \sum C_2$$

must be simultaneously a sum of blocks C_1 from P_1 and a sum of blocks C_2 from P_2. Now all sets which have this property are seen to form a complete field of sets over S. Therefore, according to Theorem 4, there exists a set of minimal blocks which are both sums of blocks from P_1 and from P_2 such that all other sets with this property are sums of these blocks. These blocks obviously define

the partition $P_1 \cup P_2$. The same argument may be used to establish the existence of the union for an arbitrary set of partitions.

Another way of obtaining the union of a set of partitions is the following. Two sets A and B in S will be said to *overlap* and we shall write $A \between B$ when their intersection $A \cdot B$ is not void. Two sets A and B shall be said to be *chain connected* by a family of subsets $\phi(A_m)$ when there exists a series of sets A_i in ϕ such that

$$A \between A_1 \between A_2 \between \cdots \between A_n \between B.$$

The family ϕ itself shall be called a *chain connected family* when any two of its sets are chain connected in it. In an arbitrary family ϕ the concept of two sets being chain connected with respect to ϕ is obviously a reflexive, symmetric and transitive relation; hence it represents an equivalence relation between the sets in the family. Thus any family of sets decomposes uniquely into maximal chain connected subfamilies such that these subfamilies have no common sets or even common elements in any of their sets.

Now let $P_1(C_1)$ and $P_2(C_2)$ be two partitions. Any block C_2 in P_2 which overlaps a block C_1 in P_1 must also belong to the same block $C_{1,2}$ of $P_1 \cup P_2$ to which C_1 belongs. By repetition of this argument one sees that $C_{1,2}$ must contain all sets C_1' and C_2' in P_1 and P_2 respectively which are chain connected with C_1 in the family formed by all blocks in the partitions P_1 and P_2. On the other hand, the maximal chain connected subfamilies in this family do define a partition of S; hence the union $P_1 \cup P_2$ has for its blocks the set of elements in the various maximal, chain connected subfamilies of the family $\{P_1(C_1), P_2(C_2)\}$. This interpretation can be extended to an arbitrary number of partitions.

This last form for the determination of the blocks in a union of partitions applies directly to the formulation in terms of equivalence relations. Let $\{E_i\}$ denote a set of equivalence relations. Their union

$$V = V\{E_i\}$$

is the equivalence relation

$$a \; V \; b$$

which holds if and only if there exists a chain of equivalences

$$a \; E_{i_1} \; a_1 , \qquad a_1 \; E_{i_2} \; a_2 , \qquad \cdots , \qquad a_{n-1} \; E_{i_n} \; b.$$

In Theorem 4 it was established that there exists a one-to-one correspondence between the partitions of S and the complete fields of sets over S. This correspondence was such that a field F corresponding to a partition P consisted of all sums of blocks in P. Now let F_1 and F_2 be two complete fields of sets over S and P_1 and P_2 the corresponding partitions. The sets common to F_1 and F_2 form the cross-cut $F_1 \cap F_2$ in the structure of complete fields over S. But the partition corresponding to $F_1 \cap F_2$ must contain both P_1 and P_2 as refinements; hence its blocks must belong both to F_1 and F_2; thus the partition corresponding to $F_1 \cap F_2$ is $P_1 \cup P_2$. Similarly the union $F_1 \cup F_2$ is the complete field gene-

rated by F_1 and F_2 ; hence its blocks must be the intersections of the blocks in P_1 and P_2 . This leads to

THEOREM 6. *There exists a dual structure isomorphism between the structure of partitions and the structure of complete fields over a set S.*

To conclude, let us mention that the union of two partitions can also be interpreted by means of a *topological graph*. Let $P(A)$ and $Q(B)$ be two partitions. The blocks A_i and B_i may be represented as points in a space. One constructs a graph with the A_i and B_i as *vertices* by joining two such vertices by an *edge* (A_i , B_i) if and only if the sets A_i and B_i overlap. This graph shall be called the *union graph* $G(P, Q)$ of the two partitions. One sees that the blocks in the union of the two partitions correspond in a one-to-one manner to the maximal connected subgraphs or components of $G(P, Q)$.

The union graph has the property that its vertices fall into two classes so that no vertex is joined by an edge to a vertex in the same class and every vertex in one class is joined by at least one edge to some vertex in the other. It is not difficult to see conversely that any graph with these properties may be considered a union graph in a suitable set.

The question arises when an arbitrary graph is a union graph. The solution follows directly from a result due to König [7]:

The necessary and sufficient condition that a graph be a union graph of two partitions is that it be a pair graph, i.e., a graph in which every cycle contains an even number of edges.

5. **The Dedekind relation.** One of the most important relations in the theory of structures is the so-called *Dedekind law*

$$A \cap (B \cup C) = B \cup (A \cap C) \qquad (A \supset B). \tag{4}$$

Sometimes it is also formulated

$$A = B \cup (A \cap C) \qquad (B \cup C \supset A \supset B). \tag{5}$$

For three equivalence relations or partitions A, B and C the Dedekind relation is usually not satisfied. The last condition in (5) states, for instance, that any block A is the sum of certain blocks B and these blocks of A are in turn contained in blocks of the union $B \cup C$. The Dedekind relation in (5) expresses that those blocks in B which make up a block in A must be chain connected by means of the family consisting of the blocks in $A \cap C$.

We shall apply this remark to determine when the Dedekind law (5) holds for *every* A containing B and contained in $B \cup C$. Let us assume first that A and B are partitions with the same blocks $A_i = B_i$ except that

$$A_1 = A_2 = B_1 + B_2 .$$

We assume further that B_1 and B_2 belong to the same block in $B \cup C$ so that A also is contained in this partition. According to the preceding formulation of the Dedekind condition (5) the two sets B_1 and B_2 must be chain connected by means of the sets in $A \cap C$. But since $A \cap C$ outside of A_1 consists only of intersections $B_i \cdot C_i$, it is only possible to obtain a connection between B_1 and B_2 in $A \cap C$ if there exists a block in C overlapping both. Conversely, it is clear that if this condition is satisfied the Dedekind law (5) must hold for all A.

It is easily seen that if (5) holds for all A then (4) also holds for all A and conversely. We shall say that the partition B is *strongly connected by means of the partition* C if to any two blocks B_1 and B_2 in B belonging to the same block in $B \cup C$ there exists some block in C overlapping both. The result which we have just derived may be expressed as

THEOREM 7. *In the structure of partitions the necessary and sufficient condition that the Dedekind law*

$$A \cap (B \cup C) = B \cup (A \cap C) \qquad (A \supset B)$$

hold for all A containing B is that the partition B be strongly connected by means of the partition C.

The condition that the partition B be strongly connected by means of the partition C can also be expressed by means of the union graph $G(B, C)$. It states that in any maximal connected component of $G(B, C)$ there will exist corresponding to any pair of vertices B_i and B_j some vertex C_k so that $G(B, C)$ contains the edges $B_i \cdot C_k$ and $B_j \cdot C_k$.

In terms of equivalence relations Theorem 7 takes the following form.

THEOREM 8. *In the structure of equivalence relations the necessary and sufficient condition that the Dedekind law*

$$E_A \cap (E_B \cup E_C) = E_B \cup (E_A \cap E_C) \qquad (E_A \supset E_B)$$

hold for all equivalence relations E_A containing E_B is that corresponding to any two elements a and b in S connected by a chain of equivalences

$$a\ E_B\ a_1\,, \qquad a_1\ E_C\ a_2\,, \qquad a_2\ E_B\ a_3\,, \qquad \cdots\,, \qquad a_r\ E_B\ b$$

there exist elements b_1 and b_2 such that

$$a\ E_B\ b_1\,, \qquad b_1\ E_C\ b_2\,, \qquad b_2\ E_B\ b.$$

We observe finally that Theorem 7 may also be stated in the form that the necessary and sufficient condition that

$$(A \cup B) \cap (A \cup C) = A \cup (C \cap (A \cup B))$$

hold for all B is that A be strongly connected by C.

We shall continue our investigations on the Dedekind law for the partition structure by proving

THEOREM 9. *The necessary and sufficient condition that the Dedekind law*

$$A \cap (B \cup C) = B \cup (A \cap C) \qquad (A \supset B) \tag{6}$$

hold for all partitions B contained in A is that there exist no circular chain of overlapping blocks

$$A_1 \between C_1 \between A_2 \between \cdots \between C_k \between A_1 \qquad (k \geq 2,\ C_1 \neq C_k) \tag{7}$$

connecting a block A_1 in A with itself.

Proof. Let us suppose first that a circular chain (7) exists. We shall then construct a partition B for which the Dedekind law (6) does not hold. We subdivide the block A_1 in A into the two blocks

$$B_1 = A_1 \cdot C_1 , \qquad \overline{B}_1 = A_1 - A_1 \cdot C_1$$

and let B consist of these two blocks and the remaining blocks of A. The existence of the chain (7) shows that the two blocks B_1 and $\overline{B}_1$ belong to the same block in $B \cup C$ and also that $A \subset B \cup C$ so that the condition (6) takes the simpler form (5). But when (5) holds the two blocks B_1 and $\overline{B}_1$ must also be chain connected by means of the blocks in B and $A \cap C$. Since A and B coincide in the blocks outside of A_1 this implies the existence of a block in C overlapping both B_1 and $\overline{B}_1$ and this contradicts the construction of these sets.

To prove the converse it is necessary to show that if no chain (7) exists the Dedekind condition (6) must be satisfied for all B contained in A. We consider a block M_1 in $B \cup C$. It consists of certain blocks in B chain connected with certain blocks in C. Since $A \supset B$ the partition $A \cap (B \cup C)$ also consists of certain sums of blocks in B. Let us assume first that for a block B_1 in B contained in some block of $A \cap (B \cup C)$ there exists no other block B_2 of B contained in M_1 for which B_1 and B_2 belong to the same A_1 in A. Then B_1 becomes a block in $A \cap B \cup C$ and it is also seen to be a block in $B \cup (A \cap C)$.

Next we assume that B_1 and B_1^* are two blocks of B contained in M_1 and both belonging to the same block A_1 . Since B_1 and B_1^* belong to the same block in $B \cup C$ there must exist a chain

$$B_1 \between C_1 \between B_2 \between C_2 \between \cdots \between C_k \between B_1^* .$$

If for every i the set B_i is contained in a corresponding A_i , this leads to the circular chain

$$A_1 \between C_1 \between A_2 \between C_2 \between \cdots \between C_k \between A_1$$

contrary to our assumption. The only way in which this contradiction can be avoided is that all A_i equal A_1 so that all blocks B_i belong to A_1. But in this case one obtains

$$B_1 \between A_1 \cdot C_1 \between B_2 \between A_1 \cdot C_2 \between \cdots \between A_1 \cdot C_k \between B_1^*;$$

hence B_1 and B_1^* are connected in $A \cap C$ and our proof is completed.

Theorem 9 can also be stated in terms of equivalence relations. The formulation by means of graphs is more interesting, however. A topological graph without cycles is called a *tree*, so that our result may be expressed as

THEOREM 10. *The necessary and sufficient condition that the Dedekind relation* (6) *hold for all B contained in A is that the union graph $G(A, C)$ be a tree.*

Since this condition is symmetric in A and C it follows that when (6) holds for all B contained in A one also has

$$C \cap (D \cup A) = D \cup (C \cap A)$$

for all D contained in C. The condition (6) may also be stated as

$$(A \cap B) \cup (A \cap C) = A \cap (C \cup (A \cap B))$$

for all B when and only when the graph $G(A, C)$ is a tree and in this case one also has

$$C \cap (A \cup (B \cap C)) = (A \cap B) \cup (A \cap C)$$

for all B.

A partition shall be called *singular* if all its blocks consist of single elements except for one block. This important type of partition occurs first in the following third result on the Dedekind relation.

THEOREM 11. *Except for the trivial cases $A = B$ and $B = 0$ the necessary and sufficient condition that the Dedekind relation*

$$A \cap (B \cup C) = B \cup (A \cap C) \tag{8}$$

hold for a fixed pair of partitions $A \supset B$ and for all C is that A be a singular partition.

Proof. It is obvious that condition (8) is satisfied for all C in the trivial cases $A = B$ or when B is the complete partition 0. We assume therefore that B has at least one block B_1 in which we can choose two elements b_1 and b_1'. Let B_1 be contained in the block A_1 of A. When $A = U$ is the universal partition it may be considered to be singular and (8) is satisfied for all B amd C. It may, therefore, be assumed also that A has a block A_2 different from A_1. We shall show first that no such block A_2 can contain two different blocks B_2 and B_2' of B. If this were the case one could choose an element b_2 in B_2 and b_2' in B_2' and define a partition C by putting

$$C_1 = \{b_1, b_2\}, \qquad C_2 = \{b_1', b_2'\}, \tag{9}$$

while all other blocks in C are single elements. Then there exists a chain

$$B_2 \between C_1 \between B_1 \between C_2 \between B_2'$$

so that B_2 and B_2' belong to the same block in $B \cup C$, consequently also in $A \cap (B \cup C)$. On the other hand, it is clear that $A \cap C = 0$ so that the right side in (8) is B and a contradiction has been obtained.

After it has been established that A and B coincide in the blocks outside of A_1 we shall show that if $A \neq B$ these blocks must consist of single elements. Since $A \neq B$ there must be at least two different blocks B_1 and B_1' of B contained in A_1. We choose b_1 in B_1 and b_1' in B_1'. If any $A_2 = B_2$ should contain two elements b_2 and b_2' we define the partition C as before by means of the two blocks (9). Again one finds a chain

$$B_1 \between C_1 \between B_2 \between C_2 \between B_1'$$

so that B_1 and B_1' belong to the same block in $B \cup C$ and $A \cap (B \cup C)$. On the other hand, one finds as before that $B \cup (A \cap C) = B$. This proves that A has to be a singular partition.

Conversely, let A have a single block A_1 with more than one element. In this case both sides of (8) are also seen to have single element blocks outside of A_1. Finally, for two blocks B_1 and B_1^* of B both contained in A_1 and belonging to the same block in $B \cup C$ one has a chain

$$B_1 \between C_1 \between B_2 \between \cdots \between C_k \between B_1^* .$$

Here all B_i must obviously belong to A_1 so that one has a chain

$$B_1 \between C_1 \cdot A_1 \between \cdots \between C_k \cdot A_1 \between B_1^*$$

and B_1 and B_1^* are connected by means of blocks in $B \cup (A \cap C)$. This concludes the proof of Theorem 11.

A *Dedekind structure* is a structure in which the Dedekind relation (4) always holds. In another paper [11] it has been shown that the necessary and sufficient condition that three elements A, B and C in an arbitrary structure generate a Dedekind structure is that the following five types of relations hold for all permutations of A, B and C:

(α) $(A \cup (B \cap C)) \cap (B \cup C) = (A \cap (B \cup C)) \cup (B \cap C)$,

(β) $(A \cup B) \cap (A \cup C) = A \cup (B \cap (A \cup C)) = A \cup (C \cap (A \cup B))$,

(β') $(A \cap B) \cup (A \cap C) = A \cap (B \cup (A \cap C)) = A \cap (C \cup (A \cap B))$,

(γ) $(A \cup B) \cap (B \cup C) \cap (C \cup A) = (A \cap (B \cup C)) \cup (B \cap (A \cup C))$ and permutations,

(γ') $(A \cap B) \cup (B \cap C) \cup (C \cap A) = (A \cup (B \cap C)) \cap (B \cup (A \cap C))$ and permutations.

Each of these relations can be analyzed in the same manner in which the Dedekind relation (4) has been studied. Such a study would seem worth while for various reasons. It would give results on the interrelations between the various conditions in the case of partitions or equivalence relations. Some of the results may possibly lead to properties of the union graphs.

6. **The distributive law.** The *distributive law* in structures

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \tag{10}$$

and certain laws related to it may be studied in the same manner as the Dedekind law in the structure of partitions.

Let us define two partitions B and C to be *associable* if the blocks in their union $B \cup C$ are blocks either in B or in C. In terms of equivalence relations this implies that if one has a chain of equivalences

$$a \; E_B \; a_1 , \qquad a_1 \; E_C \; a_2 , \qquad a_2 \; E_B \; a_3 , \qquad \cdots , \qquad a_n \; E_C \; b$$

then one of the two equivalences

$$a \; E_B \; b, \qquad a \; E_C \; b$$

must hold. In the case of associable partitions the union graph $G(B, C)$ has maximal connected components which are *stars*, in which certain vertices B_i are joined by edges to a single center C_i or conversely.

The first result on the distributive law is

THEOREM 12. *The necessary and sufficient condition that the distributive law* (10) *hold for all A is that B and C are associable.*

Proof. Condition (10) may be stated equivalently in the form that for every D contained in $B \cup C$ one has

$$D = (B \cap D) \cup (C \cap D). \tag{11}$$

To prove Theorem 12 let us suppose that a block B_1 in B overlaps at least two blocks C_1 and C_2 in C. If C_1 and C_2 are not both entirely contained in B_1 we can assume, for instance, that C_1 overlaps some block B_2. We select an element b_1 in $B_1 \cdot C_2$ and an element b_2 in $B_2 \cdot C_1$ and construct a singular partition D contained in $B \cup C$ by putting $D_1 = b_1 + b_2$. In this case

$$B \cap D = C \cap D = 0$$

so that the right side in (11) is not equal to D. Thus we have shown that any block B which overlaps more than one block of C must be a sum of blocks in C. The same argument applies to C so that B and C must be associable. Conversely, if B and C are associable the relation (11) is seen to hold for all D in $B \cup C$.

The next result is a consequence of the investigations on the Dedekind law.

THEOREM 13. *The distributive law* (10) *holds for all B in the trivial cases $C = 0$ and $A \subset C$ and otherwise only when A is a singular partition containing C.*

Proof. When the trivial cases $C = 0$ and $A \subset C$ are omitted we can assume that there exists in A some block A_1 overlapping two blocks C_1 and C_2 in C. If one of these blocks is not completely contained in A_1 we can assume, for instance, that C_2 overlaps some other block A_2 in A. We choose an element c_1 in $A_1 \cdot C_1$ and another element c_2 in $A_2 \cdot C_2$ and define B as a singular partition with $B_1 = c_1 + c_2$. Then one has $A \cap B = 0$ and $A \cap (B \cup C)$ contains the block $A_1 \cdot C_1 + A_1 \cdot C_2$ while $A \cap C$ has the blocks $A_1 \cdot C_1$ and $A_1 \cdot C_2$, so that (10) is not fulfilled. This proves that when one excludes the trivial cases one must have $A \supset C$. But in this case (10) reduces to the Dedekind relation

$$A \cap (B \cup C) = C \cup (A \cap B)$$

and it follows from Theorem 11 that this can only hold for all B when A is a singular partition.

Instead of (10) we could consider the *dual distributive law*

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \tag{12}$$

For this relation we can prove

THEOREM 14. *The dual distributive law* (12) *holds for all A only in the trivial cases $B \supset C$ or $C \supset B$.*

Proof. In order that (12) shall hold for all A it is necessary and sufficient that

$$M = (M \cup B) \cap (M \cup C) \qquad (M \supset B \cap C) \tag{13}$$

for all M containing $B \cap C$. The first step in the proof of Theorem 14 consists in showing that B and C are associable. Let us suppose that a block C_1 in C overlaps two blocks B_1 and B_2 in B. We shall prove that in this case both B_1 and B_2 are entirely contained in C_1. We assume, for instance, that the block B_1 overlaps a further block C_2. A partition M containing $B \cap C$ may then be constructed by letting M coincide with $B \cap C$ in all blocks except for the single block

$$M_1 = B_1 \cdot C_2 + B_2 \cdot C_1 .$$

It follows from this definition that $M \cup B$ contains the block $B_1 + B_2$ and $M \cup C$ contains the block $C_1 + C_2$ so that the right side of (13) is a partition in which the block

$$M_1' = B_1 \cdot C_1 + B_1 \cdot C_2 + B_2 \cdot C_1 + B_2 \cdot C_2$$

occurs. Since M_1' contains M_1 as a proper subset the two sides of (13) are not equal and we have been led to a contradiction. In the same manner it follows that if a block B_1 in B overlaps two blocks C_1 and C_2 in C, then C_1 and C_2 are entirely contained in B_1.

We can assume, therefore, that B and C are associable. If B is not contained in C nor C in B there exists a block B_1 in B containing at least two blocks C_1' and C_1'' of C and a block C_1 in C containing at least two blocks B_1' and B_1'' of B. We

construct the partition M which coincides with $B \cap C$ in all blocks except for the two larger blocks

$$M_1 = C_1' + B_1', \qquad M_2 = C_1'' + B_1''.$$

From this definition it follows that $M \cup B$ has the block $B_1' + B_1 + B_1''$ and $M \cup C$ the block $C_1' + C_1 + C_1''$; hence their cross-cut contains a block which is the intersection

$$M_1' = C_1' + C_1'' + B_1' + B_1'' = M_1 + M_2$$

of the two. The two sides of (13) are not the same in this case. This concludes the proof of Theorem 14.

Let us also deduce the following

THEOREM 15. *The distributive relation* (12) *holds for all* B *in the trivial cases* $A = 0$ *and* $A \supset C$ *and otherwise only when* C *is a singular partition containing* A.

Proof. As before, the first step in the proof is to show that A and C must be associable. Let us assume to begin with that some block A_1 in A overlaps two blocks C_1 and C_2 in C. We wish to show that C_1 and C_2 are entirely contained in A_1. If this were not the case, let C_1 overlap some other block A_2 in A. We construct the partition B which coincides with $A \cap C$ in all blocks except for the block

$$B_1 = C_1 \cdot A_2 + C_2 \cdot A_1,$$

which consists of two blocks from $A \cap C$. From this definition of B it follows that $A \cup B$ must contain the block $A_1 + A_2$ while $A \cup C$ must have a block in which $A_1 + A_2$ is entirely contained. The right side of (12) is therefore a partition in which the block $A_1 + A_2$ occurs. This leads to a contradiction since obviously $B \cap C = A \cap C$ and the left side of (12) is A. Similarly, if a block C_1 in C overlaps two blocks A_1 and A_2 in A, these two blocks must be entirely within C_1 since one would otherwise be led to the preceding case in which A_1 overlaps C_1 and C_2.

After it has been established that A and C are associable the remaining part of the proof of Theorem 15 can be based upon previous results. Since in any structure

$$(A \cup B) \cap (A \cup C) \supset A \cup (B \cap (A \cup C)) \supset A \cup (B \cap C),$$

it follows in our case from (12) that

$$(A \cup B) \cap (A \cup C) = A \cup (B \cap (A \cup C)),$$

where B is arbitrary. The result stated in Theorem 11 is now directly applicable. Except for the trivial cases where $A = 0$ or $A \supset C$, it follows that $A \cup C$ must be a singular partition. But since A and C are associable and since $A \cup C$ contains only a single block with more than one element, it is seen that one must have $C = A \cup C$; consequently C is singular and contains A.

A structure is called *distributive* when the relation (10) holds for any three elements. It can be shown [11] that three elements A, B and C in an arbitrary structure generate a distributive substructure if and only if the following three types of relations hold for all permutations of A, B and C:

$$\text{(a)}\quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

$$\text{(14) (b)}\quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$$

$$\text{(c)}\quad (A \cup B) \cap (B \cup C) \cap (C \cup A) = (A \cap B) \cup (B \cap C) \cup (C \cap A).$$

It is easily determined when the symmetric relation (14) (c) holds, for instance, for all C.

7. **The law of isomorphism.** We shall investigate to what extent the analogue of the law of isomorphism holds for equivalence relations. Let $A \supset B$ be two equivalence relations. As usual we denote the quotient structure of all equivalence relations between A and B by A/B. Furthermore, if two quotient structures A/B and A_1/B_1 are structurally isomorphic we shall write

$$A/B \cong A_1/B_1 .$$

Now let A and B be two arbitrary equivalence relations. As in the case of groups and other algebraic systems we shall study the connection between the quotient structures $D = A/A \cap B$ and $L = A \cup B/B$. In particular, we wish to determine when

$$A \cup B/B \cong A/A \cap B. \tag{15}$$

In algebraic systems the correspondences between the two quotient structures (15) are always defined by means of the so-called *regular structure correspondences* [12]

$$D \to B \cup D, \qquad A \supset D \supset A \cap B, \tag{16}$$

$$C \to A \cap C, \qquad A \cup B \supset C \supset B. \tag{17}$$

These two conditions are easily analyzed on the basis of the previous investigations on the Dedekind law. If every C in L is to be the image of some D under the correspondence (16), one sees that one must have

$$C = B \cup (A \cap C)$$

for every C between $A \cup B$ and B. We have already found that the necessary and sufficient condition for this to be the case is that B be strongly connected by A. Similarly, if every D is to be the image of some C by the correspondence (17), one must have

$$D = A \cap (D \cup B)$$

for every D between A and $A \cap B$. By the same argument which was used to prove Theorem 9, one shows that this implies that the union graph $G(A, B)$ is a tree.

These two conditions may now be combined. Through a reasoning which is particularly simple when the union graph is used, one obtains the following

THEOREM 16. *The necessary and sufficient condition that the regular structure correspondences* (16) *and* (17) *define an isomorphism*

$$A \cup B/B \cong A/A \cap B$$

for two partitions A and B is that in every block V of $A \cup B$ there shall exist a block A_1 overlapping all blocks B_j in V while all other blocks A_i in V are entirely contained in blocks B_j .

This condition can also be expressed in various other ways. In terms of the union graph it states that $G(A, B)$ consists of components in which A_1 is the center of a star of edges to the vertices B_j and from each B_j there issues another star to certain of the sets A_i , $i \geq 2$, each A_i occurring only in one of these stars.

Let us say as usual in structures that an element B is *prime under* an element A or A is *prime over* B when $A \supset B$ and the condition $A \supset X \supset B$ implies $X = A$ or $X = B$. In the case of partitions, A is prime over B when A and B coincide in all blocks except for a single block A_1 which is the sum of two blocks B_1 and B_2 from B. One verifies without difficulty the truth of

THEOREM 17. *When the partition A is prime over B and C is any other partition then either*

$$A \cup C = B \cup C$$

or $A \cup C$ is prime over $B \cup C$.

As a corollary one can state the following:

When $A \neq B$ are two partitions both prime over C the union $A \cup B$ is prime over A and B.

A *maximal chain* between two partitions $A \supset B$ is a chain

$$A \supset A_1 \supset \cdots \supset A_r \supset B,$$

where each partition is prime under the preceding partition. From the corollary of Theorem 17 one derives the following result due to Garrett Birkhoff [1; Theorems 19, 20].

In the structure of equivalence relations or partitions any two maximal chains between two elements $A \supset B$ have the same length.

The proof is analogous to the usual inductive proof for the theorem of Jordan-Hölder.

In connection with these investigations one should also mention the interesting studies by Dubreil and Mme. Jacotin-Dubreil [2], [3], [4] on chains of equivalence relations for which refinements with isomorphic quotient structures exist. These investigations are analogous to and include theories of chains of permutable subgroups in groups.

8. **Commuting equivalence relations.** Two equivalence relations E_1 and E_2 shall be said to *commute* if any two equivalences

$$a\ E_1\ c, \qquad c\ E_2\ b \tag{18}$$

imply the existence of an element d such that

$$a\ E_2\ d, \qquad d\ E_1\ b. \tag{19}$$

Conversely, in this case the equivalences (19) also imply the existence of a pair of equivalences (18).

Let us express the condition for two equivalence relations E_1 and E_2 to commute in terms of the corresponding partitions P_1 and P_2. We have seen that the union $E_1 \cup E_2$ is the equivalence relation $a\ E_1 \cup E_2\ b$ which holds if and only if a and b are connected by a chain of equivalences

$$a\ E_1\ a_1, \qquad a_1\ E_2\ a_2, \qquad \cdots, \qquad a_n\ E_2\ b.$$

When E_1 and E_2 commute this chain may be reduced to two terms

$$a\ E_1\ a_1, \qquad a_1\ E_2\ b.$$

This means that, if A and B are any two blocks of P_1 and P_2 respectively both contained in the same block V of $P_1 \cup P_2$, then A and B overlap. Conversely, if any two blocks A and B in V overlap, the two equivalences (18) and (19) will hold simultaneously and we have proved

THEOREM 18. *The necessary and sufficient condition that the two equivalence relations E_1 and E_2 commute is that the corresponding partitions P_1 and P_2 have the property that their union $P_1 \cup P_2$ consists of blocks V in which all blocks A from P_1 overlap all blocks B from P_2 and conversely.*

It is natural to say that two *partitions commute* when the corresponding equivalence relations commute. Theorem 18 shows that for the two commuting partitions A and B the blocks in the union $A \cup B$ consist of blocks A_i in A overlapping all B_j from B and conversely. Two commuting partitions are therefore strongly connected, one by the other, according to the previous definition of strong connectedness. It follows, therefore, from Theorem 7 that the Dedekind law

$$C \cap (A \cup B) = A \cup (C \cap B)$$

holds for every $C \supset A$ when A and B commute, and similarly for $C \supset B$.

From the definition of commuting partitions or equivalence relations the following properties are easily deduced:

(1) If A and B and also A and C commute, then A and $B \cup C$ commute.
(2) If $A \supset B$ then A and B commute.
(3) If A and B commute and $B_1 \supset B$, then B and $A \cap B_1$ commute.
(4) If $A_1 \supset A$ and $B_1 \supset B$ and A and B commute, then $A_1 \cap B$ and $A \cap B_1$ commute.

9. **Distributive decompositions.** Let A be an element in a structure Σ. A representation

$$A = \underset{i}{V} B_i \tag{20}$$

of A as the union of a finite or infinite number of elements will be called a *distributive decomposition* of A when for any D contained in A one has

$$D = \underset{i}{V} D_i ,$$

where D_i is contained in B_i. This may also be expressed in the form that one shall have

$$D = \underset{i}{V}(B_i \cap D).$$

When

$$A = \underset{i}{V} B_i' \tag{21}$$

is another distributive decomposition of A one can write

$$B_i' = \underset{i}{V}(B_i' \cap B_i) \tag{22}$$

so that one obtains the *refinement*

$$A = \underset{i,i}{V} (B_i' \cap B_i) \tag{23}$$

of the two distributive decompositions (20) and (21). The refinement is immediately seen to be a distributive decomposition of A.

An element A shall be said to be *distributively indecomposable* when no distributive decomposition

$$A = B_1 \cup B_2 \qquad (B_1 \neq A, B_2 \neq A)$$

exists. We shall now consider distributive decompositions (20) of A into distributively indecomposable components B_i. Such decompositions must always exist if the descending chain condition is satisfied in Σ. We assume usually that the distributive decomposition (20) is *reduced*, i.e., no B_i can be replaced by a smaller element so that the decomposition is still distributive. When the elements

B_i in (20) are distributively indecomposable the decomposition is reduced if and only if no B_i is contained in the union of the rest.

One can prove

THEOREM 19. *A reduced distributive decomposition of an element into distributively indecomposable elements is unique.*

Proof. Let (20) and (21) be two such decompositions. To prove Theorem 19 it is obviously sufficient to show that each component B_j' in (21) occurs among the components B_i in (20) and conversely. Since any B_j' is distributively indecomposable it follows that (22) cannot be an essential distributive decomposition of B_j'. Therefore, there must exist a B_i such that

$$B_j' = B_j' \cap B_i$$

or B_i contains B_j'. But similarly there must exist a B_k' containing B_i and from

$$B_k' \supset B_i \supset B_j'$$

follows

$$B_k' = B_i = B_j'.$$

A distributive decomposition (20) is said to be *direct* when each B_i is relatively prime to the union $\overline{B}_i$ of the remaining components.

$$B_i \cap \overline{B}_i = 0.$$

When two direct distributive decompositions (20) and (21) are given there exists as before a distributive refinement (23) of the two. An element A is called *direct distributively indecomposable* when no distributive decomposition (23) with $B_1 \cap B_2 = 0$ exists. One proves the following analogue of Theorem 19.

THEOREM 20. *A direct distributive decomposition of an element into direct distributively indecomposable elements is unique.*

The preceding considerations can all be dualized. Instead of the distributive decomposition (20) one obtains the *dual distributive decomposition*

$$A = \mathop{\Lambda}_i C_i \tag{24}$$

of A as the intersection of elements C_i such that for every element M containing A there exists a representation

$$M = \mathop{\Lambda}_i (C_i \cup M).$$

This condition for a dual distributive decomposition (24) can also be stated in the form that

$$L \cup A = \mathop{\Lambda}_i (L \cup C_i)$$

for an arbitrary element L.

An element A is *dually distributively indecomposable* when no dual distributive decomposition

$$A = B \cap C \qquad (B \neq A, C \neq A) \tag{25}$$

exists. The dual theorem to Theorem 19 states that, if a distributive decomposition (24) of A into dually distributively indecomposable elements C_i exists, then the decomposition is unique provided it is so reduced that no component C_i contains the intersection of the remaining ones.

The dual decomposition (24) is *direct* when for every component C_i

$$C_i \cup \overline{C}_i = U, \qquad \overline{C}_i = \underset{j \neq i}{\mathsf{A}} C_j .$$

Furthermore, an element A is *direct dually distributively indecomposable* when no dual distributive decomposition (25) exists in which $B \cup C = U$. The analogue of Theorem 20 expresses that the direct dual distributive decomposition of an element into such indecomposable components is unique.

These general remarks may be applied to the structure of all partitions of a set. We determine, first, when a partition is distributively indecomposable. It follows from Theorem 12 that a distributive decomposition

$$A = B \cup C$$

can hold only when B and C are associable partitions. From the definition of associable partitions it is seen that a partition A can always be written as a union of two associable partitions provided there exist at least two blocks in A which can be subdivided. This shows that a distributively indecomposable partition must be singular. Conversely, it is not difficult to see that a singular partition cannot be the union of two smaller associable partitions.

Any partition $P(A_i)$ is the direct union of singular partitions

$$P = \underset{i}{\mathsf{V}} P_i , \tag{26}$$

where P_i is the singular partition whose only block with more than one element is A_i . The decomposition (26) may be called the *singular decomposition* of P. Since the singular decomposition is clearly a direct distributive decomposition we conclude from the preceding that we have

THEOREM 21. *In a structure of all partitions of a set the distributively indecomposable and the direct distributively indecomposable partitions are the same, namely, the singular partitions. The reduced distributive decomposition of a partition into distributively indecomposable partitions is direct and equal to the singular decomposition.*

The dual theory yields little of interest in the case of partitions since one proves on the basis of Theorem 14 that every partition is dually distributively indecomposable.

In connection with the decomposition properties of partitions let us mention some other decompositions obtained by means of minimal and maximal partitions. It is clear that the structure of all partitions of a set has both maximal and minimal elements. The maximal partitions are those which consist of two blocks. The minimal partitions are singular partitions in which there is a single block with two elements. If these two elements of S are a and b the corresponding minimal partition may be denoted by $M_{a,b}$.

Any partition is equal to the union of the minimal elements it contains. A partition A contains the minimal partition $M_{a,b}$ if and only if a and b belong to the same block in A, or, in terms of equivalence relations, if

$$a \, E_A \, b.$$

A *basis of minimal elements* for the partition A is a set $\{M_{a,b}\}$ of minimal partitions such that

$$A = VM_{a,b}$$

and such that no subset of $\{M_{a,b}\}$ has this property. One way of obtaining a basis of minimal elements is the following. In each block A_i of A an element a_i of S is selected. The partitions M_{a_i,b_i} where b_i runs through all other elements of A_i form such a basis, when taken over all blocks A_i. This construction shows that every contraction α of the set S associated with the partition A corresponds to a basis of minimal elements for A. Obviously there are many other bases.

Similarly it is seen that every partition A is the cross-cut of maximal partitions. Any division of the set of all blocks in A into two complementary sets

$$\{A_i\}, \{A_i'\}$$

defines a maximal partition N with the two blocks

$$N_1 = \sum_i A_i, \qquad N_2 = \sum_j A_j'.$$

This partition N contains A and conversely it is clear that any maximal partition of S containing A must be of this form.

Again one can introduce a *basis of maximal elements* for A. This is a set $\{N_k\}$ of maximal partitions such that

$$A = \bigwedge_k N_k$$

is the cross-cut of these partitions while no subset of $\{N_k\}$ has this property. Such a basis may, for instance, be taken to consist of all maximal partitions N_i with the two blocks

$$N_i^{(1)} = A_i, \qquad N_i^{(2)} = S - A_i,$$

where A_i runs through all blocks of A.

Chapter 2

1. **Complements and relative complements.** Let Σ be a structure with a zero element 0 and a universal element U. A *complement* of an element A is an element $\overline{A}$ such that

$$A \cup \overline{A} = U, \qquad A \cap \overline{A} = 0.$$

As an example one may take the structure of all equivalence relations in a set S. The equivalence relation $\overline{E}$ is a complement of the equivalence relation E if any two elements a and b may be connected by a chain of equivalences

$$a\ E\ a_1\,, \qquad a_1\ \overline{E}\ a_2\,, \qquad a_2\ E\ a_3\,, \qquad \cdots\,, \qquad a_n\ \overline{E}\ b$$

and if furthermore the existence of the simultaneous equivalences

$$c\ E\ d\,, \qquad c\ \overline{E}\ d$$

always implies $c = d$.

A structure is said to be *complemented* when every element has at least one complement. Our first result is

THEOREM 1. *The structure of all equivalence relations or partitions is complemented.*

Proof. Let A be an arbitrary partition whose blocks we shall denote by A_i. In each block A_i we choose a single representative a_i and denote the total set $\{a_i\}$ by $\overline{A}_1$. The singular partition $\overline{A}$ in which $\overline{A}_1$ is the only block with more than one element is seen to be a complement of A since one verifies that

$$A \cup \overline{A} = U, \qquad A \cap \overline{A} = 0.$$

This proof even shows that one can always take the complement as a singular partition.

It may also be observed that the existence of a singular complement in the structure of partitions is equivalent to the *axiom of choice*. This remark is of some interest since it may be considered to give an algebraic formulation to the axiom of choice.

Let $A \supset B$ be two elements in a structure. We shall call an element $\overline{A}^{(B)}$ a *complement of A with respect to B* when

$$A \cup \overline{A}^{(B)} = U, \qquad A \cap \overline{A}^{(B)} = B.$$

A structure shall be said to have *relative complements* when any A has a complement with respect to any B contained in it.

Similarly, we shall call $\overline{B}^{(A)}$ a *dual complement of B with respect to A* when

$$B \cup \overline{B}^{(A)} = A, \qquad B \cap \overline{B}^{(A)} = 0$$

and the structure shall be said to have *dual relative complements* when every element B has a dual complement with respect to any A in which it is contained.

We next prove

THEOREM 2. *The structure of all equivalence relations or partitions has both relative complements and dual relative complements.*

Proof. Let A be a partition containing another partition B so that each block A_i in A contains a certain number of blocks B_j of B. We construct a relative complement $\overline{A}^{(B)}$ of A with respect to B in the following manner. In each block A_i of A we choose a single block B_i and put

$$C = \sum_i B_i .$$

The blocks in $\overline{A}^{(B)}$ shall be C and otherwise this partition shall coincide with B. One sees immediately that

$$A \cup \overline{A}^{(B)} = U, \qquad A \cap \overline{A}^{(B)} = B.$$

To construct the dual complement $\overline{B}^{(A)}$ of B with respect to A we consider a block

$$A_i = \sum_j B_j^{(i)}$$

in A. In each block $B_j^{(i)}$ contained in A_i we choose a single element $c_j^{(i)}$ and form the sets

$$C_i = \sum_j c_j^{(i)}.$$

The partition $\overline{B}^{(A)}$ shall consist of the blocks C_i and otherwise only single elements. On the basis of this construction one sees that

$$B \cup \overline{B}^{(A)} = A, \qquad B \cap \overline{B}^{(A)} = 0$$

and Theorem 2 is proved.

We shall say finally that a structure is *completely complemented* when to any three elements

$$A \supset B \supset C$$

there exists an element $\overline{B}_{A,C}$ such that

$$B \cup \overline{B}_{A,C} = A, \qquad B \cap \overline{B}_{A,C} = C.$$

Then one can state further

THEOREM 3. *The structure of all partitions is completely complemented.*

The proof of this theorem may be given by construction. It may also be based upon the following lemma which is useful for other purposes.

LEMMA. *Let A be any partition and U the universal partition of a set S. The quotient structure U/A in the structure of all partitions of S is isomorphic to the structure of partitions of the set whose elements are the blocks in A.*

The correctness of this assertion is seen immediately. Furthermore, the lemma reduces Theorem 3 to Theorem 2.

2. **Dedekind complements.** A new and stronger concept of a complement will now be introduced. A complement $\overline{A}$ of an element A in a structure Σ shall be called a *Dedekind complement* when it has the following two properties:

(1) Any B containing A has the form $B = A \cup (B \cap \overline{A})$.
(2) Any C contained in $\overline{A}$ has the form $C = \overline{A} \cap (C \cup A)$.

The condition for a Dedekind complement may also be stated equivalently by requiring that the regular structure correspondence

$$B \to B \cap \overline{A}, \qquad C \to A \cup C$$

shall establish a structure isomorphism

$$\overline{A} = \overline{A}/0 \cong U/A. \tag{1}$$

We shall not study in detail the properties of structures with Dedekind complements. It may only be mentioned that such structures must have dual relative complements and also if A/B is an arbitrary quotient structure in Σ there must exist an element A_0 such that

$$A_0 = A_0/0 \cong A/B$$

by a regular structure correspondence.

We prove

THEOREM 4. *The structure of all equivalence relations or partitions has Dedekind complements.*

This theorem may be derived directly or it may be obtained as a consequence of Theorem 16 of Chapter 1. In this theorem it was established that one has

$$A \cup B/B \cong A/A \cap B$$

by the regular structure correspondence if and only if in every block V of $A \cup B$ there exists a block A_i overlapping all blocks B_j in V while all other blocks A_i in V are entirely contained in blocks B_j.

When one assumes, in particular, that $A \cup B = U$ there exists only a single block $V = S$ and consequently one has

$$U/B \cong A/A \cap B, \qquad U = A \cup B$$

by the regular structure correspondence if and only if there exists a block A_1 in A overlapping all blocks B_i in B, while no other block A_i in A overlaps two blocks B_j. Similarly, one has

$$A \cup B/B \cong A/0, \qquad A \cap B = 0$$

by the regular structure correspondence if and only if for each block V in $A \cup B$ there exists a block A_1 in A having a single element in common with each B_i in V while all other blocks A_i contained in V are single elements.

When these remarks are applied to (1), Theorem 4 follows immediately. We also state

THEOREM 5. *The necessary and sufficient condition that $\overline{A}$ be a Dedekind complement of a partition A is that $\overline{A}$ be a singular partition.*

It may be observed that any singular partition can be considered to be a Dedekind complement of some partition. This gives a structural characterization of the singular partitions.

Relative Dedekind complements and dual relative Dedekind complements can also be defined. In the structure of partitions the existence and properties of such complements may be easily derived from previous results.

Another important problem in the theory of partitions is closely connected with the Dedekind complements. Let A be some partition and $C = \{c_i\}$ a set of elements such that there is one c_i in each block A_i of A. We shall call C a *set of representatives* for A. The choice of a set of representatives for A is obviously equivalent to the determination of a Dedekind complement for this partition. Now let A and B be two partitions. For many problems it is necessary to know when there exists a common set C of representatives for the two partitions. This requires that it be possible to enumerate the blocks A_i and B_i in A and B respectively in such a manner that A_i and B_i overlap. One obtains a set of common representatives by selecting c_i in the non-void intersection $A_i \cdot B_i$. One can obviously state

THEOREM 6. *Two partitions A and B have a set of common representatives if and only if they have a common Dedekind complement.*

Various results on the existence of sets of common representatives have been obtained. Among the most important is the following theorem of König [7], [8].

Let A and B be two partitions, both with the same finite number n of blocks. If, for $k = 1, 2, \cdots, n$, any k of the blocks A_i overlap at least k of the blocks B_i, then there exists a set of common representatives.

The theorem of König was proved first by means of the union graph. Several other proofs have been given. We shall only mention a particularly simple one given by Maak. Recently the theorem of König has been extended by Schmush-

kovitch [14]. He proves that there exists a set of common representatives when the following two conditions are satisfied:

(1) Any $k = 1, 2, \cdots$ blocks of one partition cover completely not more than k blocks of the other.

(2) Every block of one partition is covered completely by a finite number of blocks from the other.

3. **Examples from group theory.** Partitions of mathematical systems have been studied particularly in connection with the theory of groups. Among the many types of partitions or class-divisions which are of importance in group theory we shall discuss here only the so-called *co-set expansions* and even these quite briefly, mainly in order to illustrate some of the concepts already introduced.

In the following, let G denote a fixed group. Any two subgroups A and B have a cross-cut $A \cap B$ consisting of their common elements and a union $A \cup B$ which is the subgroup generated by the elements in A and B. The cross-cuts and unions may be extended to any number of subgroups so that the subgroups of a group form a complete structure.

With every subgroup A of G there are associated two partitions, the right and left co-set expansions

$$G = \sum g_i A = \sum A g_i' , \tag{2}$$

where g_i and g_i' are certain representatives in each co-set or block. We shall denote the two partitions (2) of G by P_A and ${}_AP$ respectively.

Let us now consider the combination of the two partitions P_A and P_B defined by the right co-set expansions of G with respect to two subgroups A and B. If two co-sets in these expansions have an element g in common the co-sets may be assumed to be gA and gB so that their common elements are $g(A \cap B)$. This shows that the cross-cut of the two partitions P_A and P_B is the co-set expansion $P_{A \cap B}$. Next let us consider the partition union P_A and P_B of the two co-set expansions. If a block in this union contains an element g it must contain the block gA to which g belongs in P_A and also the corresponding block gB in P_B. This shows that the same block in the union must also contain all the product sets

$$gABAB \cdots = g(A \cup B)$$

and since $g(A \cup B)$ is a sum of co-sets both in P_A and in P_B one has

$$P_A \cup P_B = P_{A \cup B} .$$

Thus we have obtained the following theorem. (This result is well known. It is stated in [1], [2], [3], [4].)

THEOREM 7. *There exists a structure isomorphism between the structure of subgroups of a group and the structure of partitions defined by the right (left) co-set expansions in the group.*

This theorem also states that the structure of all subgroups is isomorphic to a substructure of the structure of all partitions of the group.

In connection with Theorem 7 it is of interest to establish what some of the previous conditions on partitions express in the special case of co-set expansions. According to the definition, a co-set expansion P_A is *strongly connected* by means of another co-set expansion P_B when to any two co-sets in P_A which belong to the same block in $P_{A \cup B}$ there exists a block or co-sets in P_B overlapping both. Let g be an arbitrary element in some co-set with respect to $A \cup B$. Any other element g' in the same co-set has the form

$$g' = g \cdot p \qquad (p = a_1 b_1 a_2 b_2 \cdots a_k b_k),$$

where the a_i and b_i are elements in A and B respectively. When there exists a co-set with respect to B which overlaps both gA and $g'A$ there must exist two elements ga_0 and $g'a_0'$ which differ only by a right factor b_0. From

$$ga_0b_0 = g'a_0' = gpa_0'$$

it follows that the arbitrary element p in $A \cup B$ can be written in a form

$$p = aba'. \tag{3}$$

Conversely, when this is the case it is clear that the partition P_A is strongly connected by means of P_B. Furthermore, in order that an arbitrary element in $A \cup B$ have the form (3) it is sufficient to require that any product bab' have this form. This leads to the following result.

THEOREM 8. *Let A and B be two subgroups of G and P_A and P_B the partitions of G defined by the corresponding right co-set expansions. Then the partition P_A is strongly connected by means of P_B if and only if for any b_1 and b_1' in B and a_1 in A there exist elements b_2 in B and a_2 and a_2' in A such that*

$$b_1 a_1 b_1' = a_2 b_2 a_2' .$$

One can also express the condition of this theorem in the form that the union of A and B be the product ABA.

According to Theorem 7 of Chapter 1, the Dedekind relation

$$C \cap (A \cup B) = A \cup (C \cap B) \tag{4}$$

holds for any subgroup C containing A when P_A is strongly connected by means of P_B. This yields

THEOREM 9. *Let A and B be two subgroups of a group. Then the Dedekind relation* (4) *holds for all subgroups $C \supset A$ when the union of A and B is the product set*

$$A \cup B = A \cdot B \cdot A.$$

The result expressed in this theorem is of some interest since it gives a less restrictive condition for the Dedekind relation in groups than the one ordinarily

used, namely, that A and B shall be permutable subgroups. The theorem can also be verified directly. It is obviously sufficient to prove that the subgroup $A \cup (B \cap C)$ contains the subgroup $C \cap (A \cup B)$ because the converse is trivial. Let c be an element in $C \cap (A \cup B)$. Since P_A is connected by means of P_B one can write c in the form

$$c = a_1 b_1 a_2 ,$$

or

$$b_1 = a_1^{-1} \cdot c a_2^{-1} \subset C \cap B;$$

hence c belongs to $A \cup (B \cap C)$.

Two partitions P_A and P_B were said to *commute* if in every block of $P_{A \cup B}$ any block of P_A overlaps every block of P_B and conversely. If p is an arbitrary element in $A \cup B$, two arbitrary elements in the same block of $P_{A \cup B}$ may be assumed to have the form g and $g \cdot p$ and belong to the co-sets gA and gpA in P_A. If there exists a co-set $g \cdot B$ in P_B containing g and overlapping gpA there must exist an element b in B such that

$$g \cdot b = g \cdot p \cdot a;$$

hence any p in $A \cup B$ has the form

$$p = a_1 \cdot b_1 .$$

In the usual terminology of group theory, the two subgroups A and B are said to be *permutable* in this case so that we have [4; Theorem 2]

THEOREM 10. *The necessary and sufficient condition that the two partitions P_A and P_B commute is that the subgroups A and B be permutable.*

We shall now consider the combination of right and left co-set expansions. Let A and B be subgroups as before and ${}_AP$ the left co-set expansion with respect to A and P_B the right co-set expansion with respect to B. In order to determine the cross-cut of these two partitions let Ag and gB be two of their blocks containing a common element g. The other common elements of the two co-sets must be of the form

$$d = a \cdot g = g \cdot b;$$

hence a belongs to

$$A \cap gBg^{-1}$$

and b to

$$B \cap g^{-1}Ag.$$

The blocks in the intersection of ${}_AP$ and P_B are therefore the left and right co-sets

$$g \cdot (B \cap g^{-1}Ag) = (A \cap gBg^{-1}) \cdot g.$$

When g is an element in a block of the union ${}_AP \cup P_B$ one sees that the same block must contain all the elements in the *double co-set* AgB. Conversely, this double co-set is a sum of left co-sets of A and right co-sets of B so that we have obtained

THEOREM 11. *Let ${}_AP$ and P_B be the left and right co-set expansions of a group G with respect to the subgroups A and B respectively. The cross-cut ${}_AP \cap P_B$ of the two partitions consists of the blocks*

$$g(B \cap g^{-1}Ag) = (A \cap gBg^{-1})g$$

while the union is a double co-set expansion

$$G = \sum AgB.$$

From the point of view of the theory of partitions there exists the following interesting relation between the right and left co-set expansions.

THEOREM 12. *The partition ${}_AP$ of a left co-set expansion always commutes with the partition P_B of a right co-set expansion and conversely.*

Proof. We have seen that the blocks in the union ${}_AP \cup P_B$ are double co-sets AgB. Any element g_1 in this double co-set generates it in the sense that one can obtain any other element g_2 in AgB by multiplying g_1 on the left by an element a and on the right by an element b. But this statement is seen to be the same as saying that any block in ${}_AP$ overlaps any block in P_B in AgB and conversely.

When a block in ${}_AP \cup P_B$ contains the same finite or infinite number of blocks of ${}_AP$ and P_B one can establish a one-to-one correspondence $A_\alpha \rightleftarrows B_\alpha$ between these two co-sets. A common representative element can then be chosen for A_α and B_α from the non-void intersection $A_\alpha \cdot B_\alpha$, so that one can write

$$AgB = \sum Ag_\alpha = \sum g_\alpha \cdot B$$

with the same right and left multipliers. This remark applies particularly to the case where A and B are finite groups with the same order, since for such a group AgB is also a finite set. We have deduced therefore the following theorem of G. A. Miller [9], [10].

Let A and B be finite subgroups of the same order in a group G. Then there exists a set of common representatives g_α for the left and right co-set expansions of G with respect to A and B so that

$$G = \sum Ag_\alpha = \sum g_\alpha B.$$

This theorem can be extended immediately to the case where A and B are subgroups with *finite indices* because in this case, as it is well known, G can be represented homomorphically as a finite substitution group.

The theorem of Miller is not true in general for infinite subgroups, not even in the important special case $A = B$. This follows from an example given recently by Shü [15].

Chapter 3

1. **Correspondences and inverse correspondences.** The theory of partitions is closely connected with the theory of *correspondences of sets*. We shall now study the partitions from this point of view. Let α be some *correspondence* which takes each element a of the set S into some uniquely defined element a^α. In general, several elements a may correspond to the same element a^α and the set of images S^α is usually a proper subset of S.

To any correspondence α one can introduce an *inverse correspondence* α^{-1}. For any subset A of S we shall denote by $A^{\alpha^{-1}}$ the set of all elements in S whose image under α lies in A. This may also be stated thus: $B = A^{\alpha^{-1}}$ is the largest subset of S such that

$$B^\alpha = (A^{\alpha^{-1}})^\alpha = A \cdot S^\alpha.$$

The two correspondences α and α^{-1} determine each other uniquely. While α is a many-to-one correspondence the inverse correspondence may be termed a one-to-many correspondence.

The correspondence α is determined when it is known which set of elements

$$C_a = a^{\alpha^{-1}} \tag{1}$$

corresponds to a given element a. Let us notice that if A_1 and A_2 are disjoint subsets of S the two sets $A_1^{\alpha^{-1}}$ and $A_2^{\alpha^{-1}}$ are also disjoint. When this remark is applied to the sets (1) it follows that the sets (1) are disjoint and form the blocks in a partition of S. Thus there exists for every correspondence α a unique *associated partition*

$$P_\alpha = P_\alpha(C_a).$$

The associated partition P_α is the complete partition of S if and only if α is a one-to-one correspondence of S to a subset.

Two correspondences α and β of the set S to subsets S^α and S^β shall be said to be *conformal* when they are associated with the same partition. Two conformal correspondences α and β each associate a single element with each block C in the corresponding partition. This leads to

THEOREM 1. *The necessary and sufficient condition that the two correspondences α and β be conformal is that*

$$\beta = \tau_0 \cdot \alpha,$$

where τ_0 is a one-to-one correspondence between S^α and S^β.

Let α, β and γ be three correspondences of the set S to subsets such that

$$\alpha = \beta \cdot \gamma, \qquad a^\alpha = (a^\gamma)^\beta.$$

In this case we shall say that α contains β and γ as *left* and *right factors* respectively. Since $a_1^\alpha \neq a_2^\alpha$ must imply $a_1^\gamma \neq a_2^\gamma$ it follows that one can have $a^\gamma = b^\gamma$ only when $a^\alpha = b^\alpha$. This shows that when γ is a right factor of α the associated partition P_γ must be contained in the partition P_α. Conversely we can show that this is a sufficient condition for γ to be a right factor of α. Let us suppose namely that $P_\alpha \supset P_\gamma$ for two correspondences α and γ. Then each block A_i in P_α is the sum of certain blocks $C_{i,j}$ in P_γ

$$A_i = \sum C_{i,j} .$$

Let us write further

$$A_i^\alpha = a_i , \qquad C_{i,j}^\gamma = c_{i,j}$$

so that

$$A_i^\gamma = \sum_\gamma c_{i,j} .$$

A correspondence β such that $\alpha = \beta \cdot \gamma$ can now be defined by putting

$$c_{i,j}^\beta = a_i$$

and letting β be defined arbitrarily for all other elements in S. This proves

THEOREM 2. *Let α and γ be two correspondences of a set S to subsets. Then γ is a right factor of α if and only if the partition P_α associated with α contains the partition P_γ associated with γ.*

One can also ask analogously when a correspondence is a left factor of another correspondence α. In this case one concludes from

$$S^\alpha = (S^\gamma)^\beta$$

that $S^\beta \supset S^\alpha$ and this condition can also be shown to be sufficient for a left factor. To demonstrate this, let a be any element in S^α and

$$A_a = a^{\alpha^{-1}}, \qquad B_a = a^{\beta^{-1}} .$$

In each B_a we choose a single element b_a and define the correspondence γ conformal to α by putting

$$A_a^\gamma = b_a .$$

It follows immediately that

$$(A_a^\gamma)^\beta = b_a^\beta = a$$

so that $\alpha = \beta \cdot \gamma$. We have obtained

THEOREM 3. *Let α and β be correspondences of a set S to subsets S_α and S_β. The necessary and sufficient condition that β be a left factor of α is that $S^\beta \supset S^\alpha$. When in this case $\alpha = \beta \cdot \gamma$ the correspondence γ may be taken to be conformal with α.*

2. **Contractions.** Let $P = P(C)$ be a partition of the set S. To any such partition there exists a correspondence γ having $P = P_\gamma$ for its associated partition. This correspondence can, for instance, be chosen in the following particular manner. In each block C we choose a single representative element c_C. A correspondence γ is obtained by letting each element in C correspond to c_C, hence

$$C_\gamma = c_C \in C.$$

A correspondence of this type shall be called a *contraction* of the set S. From the definition of a contraction and from Theorem 1 follows directly

THEOREM 4. *To any partition P there exist contractions with P as its associated partition. Any correspondence α is conformal with a partition γ and can be written in the form*

$$\alpha = \tau_0 \cdot \gamma,$$

where γ is a one-to-one correspondence between the subsets of S.

The contractions can also be characterized in another manner. Let us say that a correspondence γ of a set S to a subset S^γ is *idempotent* when $\gamma^2 = \gamma$. Then we can show

THEOREM 5. *Any contraction is idempotent and conversely any idempotent correspondence of S to a subset is a contraction.*

Proof. The first part of the theorem follows immediately from the definition of a contraction. To prove the converse let γ be an idempotent correspondence and a some element in S^γ. In this case the set $C_a = a^{\gamma^{-1}}$ is not void and one finds

$$a = (a^{\gamma^{-1}})^\gamma = (a^{\gamma^{-1}})^{\gamma \cdot \gamma} = a^\gamma$$

so that

$$C_a^\gamma = a \in C_a .$$

The product of two contractions is usually not a contraction. We shall now determine the conditions for this to be the case. We consider two contractions α and β of the set S with the corresponding partitions P_α and P_β. Let A_0 be some fixed block in P_α with the representative element a_0 under α. We assume that A_0 contains the representative elements $b_{0,i}$ under β of certain blocks $B_{0,i}$ in P_β. When

$$C_0 = \sum_i B_{0,i} \tag{2}$$

is the sum of these blocks one obtains

$$C_0^{\alpha\beta} = (\sum_i B_{0,i})^{\alpha\beta} = \sum_i b_{0,i}^{\alpha} = a_0$$

so that C_0 is the set of all elements having the image a_0 under the correspondence $\alpha\beta$; hence C_0 is a block in $P_{\alpha\beta}$. This shows that $\alpha\beta$ is a contraction if and only if a_0 belongs to C_0 and therefore also to some $B_{0,i}$ having its image $b_{0,i}$ under β in A_0.

In order to simplify the formulation of this and the following results, let us say that for two contractions α and β the two blocks A in P_α and B in P_β are *paired* if their representative elements a and b under α and β respectively both belong to the intersection $A \cdot B$. We can then state the preceding result as

THEOREM 6. *Let α and β be two contractions of the set S with the associated partitions $P_\alpha(A)$ and $P_\beta(B)$. The product $\alpha\beta$ is a contraction if and only if every block A_0 in P_α containing at least one representative element b under β is paired with some block B_0.*

This result may be stated in somewhat different forms. For instance, corresponding to any representative element b under β there shall exist a particular representative element b_0 such that

$$b^\alpha = b_0^\alpha = a_0 , \qquad a_0^\beta = b_0 ,$$

where a_0 is a representative under α. One can also say that corresponding to every b there will exist a b_0 such that

$$b^\alpha = b_0^\alpha , \qquad b_0^{\beta\alpha} = b_0 .$$

It can be shown easily from Theorem 6 that to a given contraction β one can always determine another contraction α with an arbitrary prescribed associated partition $P = P_\alpha$ such that $\alpha\beta$ is a contraction.

One deduces directly from Theorem 6 the following

THEOREM 7. *Let α and β be two contractions. The necessary and sufficient condition that $\alpha\beta$ and $\beta\alpha$ both be contractions is that the blocks in $P_\alpha(A)$ and $P_\beta(B)$ either be paired, or that any non-paired block A shall not contain any representatives from β and have its own representative under α in a paired block B and similarly for any non-paired block B.*

A consequence of Theorem 7 is

THEOREM 8. *For any two partitions there exist associated contractions α and β such that $\alpha\beta$ and $\beta\alpha$ are also contractions.*

Proof. Let $P_1(A)$ and $P_2(B)$ be two partitions. We shall prove first that the blocks in any two partitions can always be paired in such a way that any two paired blocks A and B overlap while no non-paired block B can overlap any non-paired A. To prove this statement let us assume that the blocks $P_1(A)$ are well-ordered. To any A_ω we pair some block $B_{\omega'}$ which overlaps A_ω and which has not been paired with any previous block A. If no $B_{\omega'}$ of this kind exists, A_ω remains non-paired and it is obvious that it does not overlap any non-paired B.

After this pairing of blocks has been completed, no remaining non-paired B can overlap a non-paired A because otherwise the two could have been paired.

After this auxiliary result has been established one can determine the contractions α and β such that for a pair of sets A and B a common representative under α and β is selected in the intersection $A \cdot B$. For the non-paired blocks A we choose the representative under α in one of the paired blocks B which it overlaps and similarly for a non-paired block B. The conditions of Theorem 7 are seen to be satisfied and Theorem 8 is proved.

Let us finally determine the conditions for two contractions α and β to commute.

$$\alpha\beta = \beta\alpha.$$

When this condition is satisfied the product $\alpha\beta$ is obviously idempotent so that we have a special case of the situation considered previously in Theorem 7. According to Theorem 2 one must have

$$P_{\alpha\beta} \supset P_\alpha , \qquad P_{\alpha\beta} \supset P_\beta$$

so that

$$P_{\alpha\beta} \supset P_\alpha \cup P_\beta .$$

On the other hand, it is clear that by the successive application of the contraction α and the contraction β to the elements in a block C of $P_\alpha \cup P_\beta$ one obtains a single image also contained in C so that one actually has

$$P_{\alpha\beta} = P_\alpha \cup P_\beta .$$

To determine the further necessary conditions for α and β to commute, let us turn to the expression (2) for the blocks in the partition $P_{\alpha\beta}$. One sees that all the blocks $B_{0,i}$ in C_0 have their representatives under β in the same block A_0. This means that when α and β commute there must exist in each block C of $P_\alpha \cup P_\beta$ some block A_0 overlapping all blocks B in C and containing all representatives under β of these blocks B. For reasons of symmetry there must also exist a block B_0 overlapping all blocks A in C and containing all their representatives under α. Furthermore, if a_0 is the representative of A_0 under α and b_0 the representative of B_0 under β, then one must have

$$C_0^{\alpha\beta} = a_0 = C_0^{\beta\alpha} = b_0$$

so that A_0 and B_0 have the same representative. Conversely, let us assume that in any block C of $P_\alpha \cup P_\beta$ there exists a single block A_0 containing the representatives under β of all the blocks B in C and similarly a single block B_0 containing the representatives under α of all the blocks A in C. Then if A_0 and B_0 have the same representative under α and β respectively, it is clear that α and β commute. This leads to

THEOREM 9. *Let α and β be two contractions with the associated partitions P_α and P_β. The necessary and sufficient condition that α and β commute is that*

$$P_{\alpha\beta} = P_\alpha \cup P_\beta$$

and in any block C of $P_{\alpha\beta}$ there exist blocks A_0 and B_0 of P_α and P_β respectively with a common representative

$$c = A_0^\alpha = B_0^\beta$$

and such that all blocks B of P_β in C have their representative in A_0 and all blocks A of P_α have their representative in B_0.

An immediate consequence of this theorem is

THEOREM 10. *Let $P_1(A)$ and $P_2(B)$ be two partitions. The necessary and sufficient condition that one can define commuting contractions α and β associated with them is that in every block C in $P_1 \cup P_2$ there exist a block A_0 overlapping all blocks B and a block B_0 overlapping all A.*

In this case the two partitions are obviously strongly connected according to the definition of this concept. Two commuting equivalence relations or partitions P_1 and P_2 had the characteristic property that *any* A overlapped *any* B in the same block C of the union, and similarly any B overlapped any A. Thus one can associate commuting contractions with commuting partitions, but the converse is not necessarily true.

3. **Automorphisms and endomorphisms of equivalence relations.** Let α be a one-to-one correspondence or the *transformation* of the set S into itself. We shall say that α is an *automorphism* of the equivalence relation E when any equivalence

$$a \, E \, b$$

implies

$$a^\alpha \, E \, b^\alpha, \qquad a^{\alpha^{-1}} \, E \, b^{\alpha^{-1}}.$$

This means that two elements a and b which belong to the same block A in the partition P_E defined by E cannot be transformed into elements belonging to two different blocks by α or by α^{-1}. This remark leads to

THEOREM 11. *A transformation α is an automorphism of the equivalence relation E if and only if every block A in the associated partition P_E is transformed into other blocks A^α and $A^{\alpha^{-1}}$ of P_E by α and α^{-1}.*

The set of all automorphisms form a group $\mathfrak{A}(E)$, the *group of automorphisms* of the equivalence relation E. In order to determine this group in explicit form let us first divide the blocks in the partition P_E associated with E into subfamilies $P_i = P_i(C_i)$ in such a manner that each subfamily P_i consists of those blocks in

P_E whose number of elements has the same cardinal number γ_i . Let us also write $S_i = \sum C_i$ for the sum of the sets in one of these subfamilies. By an automorphism α of E a block C_i must obviously always be transformed into another block C_i^α in the same subfamily P_i . This shows that any automorphism α can be written as a product

$$\alpha = \prod_i \alpha_i ,$$

where α_i transforms the set S_i into itself and leaves all elements outside of S_i invariant. An equivalence relation shall be called *homogeneous* if all its blocks contain the same cardinal number γ of elements. The group of automorphisms of E is therefore a direct product

$$\mathfrak{A}(E) = \prod_i \mathfrak{A}(E_i) \tag{3}$$

of groups of automorphisms $\mathfrak{A}(E_i)$ where E_i is the homogeneous equivalence relation defined in the set S_i by the blocks C_i of the family P_i .

Through the expression (3) the determination of the group of automorphisms has been reduced to the case of a homogeneous equivalence relation E. We now suppose therefore that the partition $P = P(C_i)$ associated with E has κ blocks, each containing γ elements, where κ and γ may be arbitrary cardinal numbers. Let $\alpha_{i,j}$ be some one-to-one correspondence between the elements in the two blocks C_i and C_j and let us denote this fixed correspondence simply by $C_i \to C_j$. Furthermore, let τ_i be arbitrary transformations of the set C_i into itself. It is then clear that any automorphism of E can be written formally as a generalized permutation

$$\begin{pmatrix} C_1 , C_2 , \cdots \\ \tau_{i_1} : C_{i_1} , \tau_{i_2} : C_{i_2} , \cdots \end{pmatrix}.$$

The group consisting of all generalized permutations

$$\begin{pmatrix} x_1 , x_2 , \cdots , x_n \\ \tau_1 x_{i_1} , \tau_2 x_{i_2} , \cdots , \tau_n x_{i_n} \end{pmatrix},$$

where each variable is transformed into some other variable multiplied by a factor τ belonging to a group T, is called a *complete monomial group* over T. It may be denoted by $\Sigma_n(T)$ where T is called the *coefficient group*. Without going into the details it is not difficult to see that the group of automorphisms of a homogeneous equivalence relation E is a complete monomial group $\Sigma_\kappa(G_\gamma)$ where the coefficient group G_γ is the symmetric group on γ elements. (A fairly complete theory of monomial groups can be found in [13].) We can state, therefore,

THEOREM 12. *Let E be an equivalence relation, and let the associated partition P_E for each i contain κ_i blocks with γ_i elements, where κ_i and γ_i may be arbitrary cardinal numbers. Then the group of automorphisms of E is a direct product*

$$\mathfrak{A}(E) = \prod_i \mathfrak{A}_i(E_i)$$

where each factor is a complete monomial group

$$\mathfrak{A}_i(E_i) = \Sigma_{\kappa_i}(G_{\gamma_i}),$$

where the coefficient group G_i is the symmetric group on γ_i elements.

A correspondence α of the set S to a subset S^α will be called an *endomorphism* of the equivalence relation E if any equivalence

$$a \; E \; b$$

implies

$$a^\alpha \; E \; b^\alpha.$$

The product of two endomorphisms is another endomorphism so that the endomorphisms form a multiplicative system of correspondences, the *endomorph* $\mathfrak{E}(E)$ of E. An endomorphism can obviously be characterized by

THEOREM 13. *A correspondence α is an endomorphism of the equivalence relation E if any block C in the associated partition P_E corresponds under α to some subset of some other block C' of the same partition.*

All endomorphisms of E can therefore be obtained by first selecting a certain number of fixed but arbitrary blocks C' in P_E. One constructs α by letting each C in P_E correspond in some manner to a subset of one of the blocks C'.

4. **Representation of associative systems by correspondences.** Let $\mathfrak{M}$ denote some multiplicative system in which the multiplication satisfies the associative law

$$(ab)c = a(bc).$$

Any such system can be represented by means of correspondences of the set $\mathfrak{M}$ by associating with each element a the correspondence γ_a defined by

$$x \to ax = x^{\gamma_a}. \tag{4}$$

The association

$$a \to \gamma_a$$

shall be called the *regular representation* of $\mathfrak{M}$. The regular representation is a homomorphism of $\mathfrak{M}$ since

$$x^{\gamma_a \cdot \gamma_b} = (x^{\gamma_b})^{\gamma_a} = (bx)^{\gamma_a} = abx = x^{\gamma_{ab}}$$

so that

$$\gamma_a \cdot \gamma_b = \gamma_{ab} .$$

The regular representation is an isomorphic representation of $\mathfrak{M}$ provided $a \neq b$ implies $\gamma_a \neq \gamma_b$. This requires that if $a \neq b$ there exist at least one element x such that

$$ax \neq bx.$$

This condition is always satisfied when $\mathfrak{M}$ has a unit element. If this is not the case, a unit element e can be adjoined to $\mathfrak{M}$ and for the enlarged system there exists an isomorphic regular representation. This proves that every associative system can be represented isomorphically by means of correspondences.

These considerations may be applied to systems with a so-called *algebraic set operation*. Such a system $\mathfrak{M}$ possesses an operation $a \cup b$, for instance, called a *union*, which has the following three properties of being:

(1) idempotent: $a \cup a = a$,
(2) commutative: $a \cup b = b \cup a$,
(3) associative: $(a \cup b) \cup c = a \cup (b \cup c)$.

Any such system may be considered to be a partially ordered set when one puts $a \supset b$ whenever $a \cup b = a$. Its regular representation is isomorphic since

$$a \cup x = b \cup x$$

for all x in $\mathfrak{M}$ implies

$$a = a \cup a = b \cup a = b \cup b = b.$$

It follows from the preceding statements that the correspondences in the representation are all contractions, and since they commute they satisfy the special conditions of Theorem 9.

This procedure may also be applied to represent a structure by means of contractions. In this case one obtains one representation by means of the union and another by means of the cross-cut. These two are not the same. Let us consider, for instance, the representation defined by means of the union. Since a contraction is associated with a unique partition, every element a in the structure Σ also becomes associated with a partition P_a. Conversely one verifies that a is uniquely defined by P_a. If, namely, $a \neq b$ then $P_a \neq P_b$ since one finds that the block in P_a to which the cross-cut $a \cap b$ belongs consists of the elements contained in a while the corresponding block in P_b consists of the elements contained in b. The correspondence $a \rightleftarrows P_a$ may be called the *union partition representation* of Σ. From Theorem 9 one concludes that one has

$$a \cup b \rightleftarrows P_a \cup P_b .$$

Thus one has imbedded the structure Σ in the structure of partitions of Σ in such a manner that unions and hence also order are preserved. Usually this

representation $a \rightleftarrows P_a$ is not a structure isomorphism. A block in the partition $P_a \cap P_b$ consists of all x such that

$$a \cup x_0 = a \cup x, \qquad b \cup x_0 = b \cup x \tag{5}$$

for some fixed x_0. On the other hand, the partition $P_{a \cap b}$ consists of all blocks of all elements x such that

$$(a \cap b) \cup x_0 = (a \cap b) \cup x. \tag{6}$$

From (6) one can always conclude that the relations (5) hold so that one has

$$P_{a \cap b} \subset P_a \cap P_b .$$

But when Σ is a distributive structure, (6) is a consequence of (5) so that we can state

THEOREM 14. *In a distributive structure Σ the correspondence $a \rightleftarrows P_a$ between an element a of Σ and the union partition P_a is a structure isomorphism.*

Chapter 4

1. **Geometry of partitions.** We shall consider the problem of giving an axiomatic characterization of the structure of partitions and equivalence relations. This part of the theory of equivalence relations can be formulated in different ways. We shall prefer here a geometric form which brings the results into an interesting relation with other more familiar geometric theories.

In the following, we shall study a certain basic set G which we shall call the *geometric system*. The elements in G shall be called *points*. With every pair of different points P_1 and P_2 in G there will be associated a unique set $l(P_1 , P_2)$ of elements in G. This set $l(P_1 , P_2)$ we shall call the *line* defined by P_1 and P_2.

Certain axioms will now be imposed upon the properties of these lines and the final goal is to show that these axioms suffice to identify our geometric system with the structure of all partitions of some set S.

Our first axiom is

AXIOM 1. *The line $l(P_1 , P_2)$ contains the points P_1 and P_2 and at most one further point P_3.*

Two points shall be called *related* when the line they define contains three points. Otherwise they are *unrelated*. Three points P_1, P_2 and P_3 on a line shall be said to be *collinear*. When P_3 is the third point on a line $l(P_1 , P_2)$ we shall write

$$P_3 = \beta(P_1 , P_2)$$

and call P_3 the *bond* between the related points P_1 and P_2.

The next axiom is

AXIOM 2. *Any two points on a line define the same line.*

This axiom states that for collinear points P_1 , P_2 and P_3 one has

$$l(P_1 , P_2) = l(P_2 , P_3) = l(P_3 , P_1).$$

Thus if (P_1 , P_2) is a related pair of points the pairs (P_2 , P_3) and (P_3 , P_1) are also related. Furthermore, one has

$$P_3 = \beta(P_1 , P_2), P_1 = \beta(P_2 , P_3), P_2 = \beta(P_3 , P_1)$$

and any one of these three relations implies the two others.

An immediate consequence of Axiom 2 is:

Two different lines have at most one point in common.

Before we proceed further into this quasi-geometrical theory it might be well to clarify its relation to the structure of all partitions of a set. This will also elucidate the reasons for introducing the subsequent axioms. The minimal partitions or points P among the partitions of a set S are the singular partitions with one block containing two elements of S. When these two elements are a and b we shall write $P = P_{a,b}$. We introduce the line defined by two minimal partitions P and Q as the set of minimal partitions contained in the union $P \cup Q$. One sees that two minimal partitions $P_{a,b}$ and $P_{c,d}$ are related if and only if c or d is equal to one of the elements a or b. Thus $P_{a,b}$ and $P_{b,c}$ will be the general form for two related minimal partitions and their third collinear minimal element is

$$P_{a,c} = \beta(P_{a,b} , P_{b,c}).$$

After these remarks we return to the general case and assume further

AXIOM 3. *Let P_1 , P_2 and P_3 be collinear points and R a point not on this line such that the line $l(R, P_1)$ contains three points. Then one and only one of the lines $l(R, P_2)$ and $l(R, P_3)$ will contain three points.*

This condition may also be stated in the form that if R is related to one of the three collinear points it is related to one and only one of the others.

We shall now introduce a concept which is fundamental for the following theory. Let P_1 , P_2 , P_3 be collinear points. All points R different from P_3 such that the lines $l(R, P_1)$ and $l(R, P_2)$ contain three points will be said to constitute a *star*. We shall denote the star by $\sigma(P_1 , P_2)$ and by special definition we shall include the points P_1 and P_2 (but not P_3) in it. According to Axiom 3 a point R different from P_1 , P_2 and P_3 can at most belong to one of the stars

$$\sigma(P_1 , P_2), \sigma(P_2 , P_3), \sigma(P_3 , P_1).$$

Consequently any two of these stars have only one point in common; for instance, $\sigma(P_1, P_2)$ and $\sigma(P_2, P_3)$ have P_2 in common.

In order to prove a fundamental property of the stars we must assume further

AXIOM 4. *Let P_1, P_2, P_3 be collinear points and Q and R points outside this line such that the lines*

$$l(Q, P_1),\ l(Q, P_2),\ l(R, P_1),\ l(R, P_2)$$

contain three points. Then the line $l(Q, R)$ will contain three points.

This axiom states that if Q and R are two points belonging to the same star then they are related.

On the basis of Axiom 4 one proves:

When R is a point belonging to a star $\sigma(P_1, P_2)$ then

$$\sigma(P_1, P_2) = \sigma(R, P_1). \tag{1}$$

Proof. First it is clear that P_1 and R belong to both these two stars. Furthermore P_2 is also contained in the second star (1) since P_2 is related to both R and P_1 and one cannot have $P_2 = \beta(R, P_1)$ since this would imply $R = \beta(P_1, P_2)$, a case excluded by the definition of a star. Next let Q be some other point contained in the star $\sigma(P_1, P_2)$. According to Axiom 4 the point Q is related to both P_1 and R; hence Q belongs to the star $\sigma(R, P_1)$ provided one does not have $Q = \beta(R, P_1)$. But Axiom 3 shows that this is impossible since the point P_2 would be related to the three collinear points P_1, R, Q. In the same manner one shows that every point Q in $\sigma(R, P_1)$ belongs to $\sigma(P_1, P_2)$ and the identity of the two stars (1) is shown.

The preceding result implies the following important property.

Any two points in a star determine the same star.

Proof. Let $\sigma(P_1, P_2)$ be the given star and Q and R points contained in it. According to (1) one concludes

$$\sigma(P_1, P_2) = \sigma(P_1, R) = \sigma(Q, R).$$

We state further:

Any two different stars can have at most one point in common.

Proof. If they had two points in common they would be identical.

A point cannot belong to more than two stars.

Proof. Let P_1 be the given point. If P_1 belongs to a star there must exist three collinear points P_1, P_2, P_3 and P_1 belongs to the two stars $\sigma(P_1, P_2)$ and $\sigma(P_1, P_3)$. To see that it cannot belong to any other star we observe first that

this star could be represented in the form $\sigma(P_1 , R)$. The point R would then be related to P_1, hence also to one of the points P_2 and P_3 according to Axiom 3. But if R is related, for instance, to P_2 the preceding results give

$$\sigma(P_1 , R) = \sigma(P_1 , P_2).$$

Two points are related if and only if they belong to the same star.

Proof. If they are unrelated they cannot both belong to the same star and if they are related they define a star to which they both belong. They cannot then both belong to some other star.

Two stars shall be said to be *connected* when they have a point in common. We are able to prove on the basis of the previous axioms the following transitive property of connectedness.

When the star σ_1 is connected with the star σ_2 by the point $P_{1,2}$ and σ_2 with the star σ_3 by the point $P_{2,3}$, then σ_1 is connected with σ_3 by the point

$$P_{1,3} = \beta(P_{1,2} , P_{2,3}). \tag{2}$$

Proof. Since $P_{1,2}$ and $P_{2,3}$ both belong to σ_2 they are related; hence the collinear point $P_{1,3}$ defined by (2) exists. We shall have to show that $P_{1,3}$ belongs both to σ_1 and σ_3. In order to show that $P_{1,3}$ belongs to σ_1, let Q denote an arbitrary point in σ_1 different from $P_{1,2}$. Since Q is related to $P_{1,2}$ it must be related to one of the points $P_{2,3}$ and $P_{1,3}$ according to Axiom 3. But if Q were related to $P_{2,3}$ it would belong to the star

$$\sigma_2 = \sigma(P_{1,2} , P_{2,3})$$

and one would have $Q = P_{1,2}$ contrary to assumption. Therefore Q is related to $P_{1,3}$ and so it belongs to the star $\sigma(P_{1,2} , P_{1,3})$. But then $P_{1,3}$ belongs to the same star

$$\sigma(P_{1,2} , P_{1,3}) = \sigma(P_{1,2} , Q) = \sigma_1 .$$

Similarly one proves that $P_{1,3}$ belongs to σ_3.

We now introduce the final axiom.

AXIOM 5. *For any two points P and Q such that the line $l(P, Q)$ contains only two points there will exist a third point R so that both the lines*

$$l(P, R), \qquad l(R, Q)$$

contain three points.

This axiom states in other terms that for any two unrelated points P and Q there will exist a point R related to both of them.

When Axiom 5 is assumed it is possible to prove:

Any two stars are connected.

Proof. Let σ_1 and σ_2 be two stars and P and Q a point in each of them. According to Axiom 5 there exists a point R related to both P and Q. But then the stars

$$\sigma_1 , \qquad \sigma(P, R), \qquad \sigma(R, Q), \qquad \sigma_2$$

form a series in which each star is connected with the preceding one; hence σ_1 and σ_2 are also connected by the transitive property of connectedness.

When we exclude the trivial case where there is only one point, it follows from Axiom 5 that to every point P there exists at least one point related to it. This means that every point belongs to at least one star; hence we can state on the basis of a previous result:

Any point belongs to exactly two stars which it connects.

We may therefore characterize any point $P = P(\alpha, \beta)$ uniquely by the two stars α and β which it connects. Since we have observed that two points P and Q are related if and only if they belong to the same star, it follows that $P(\alpha, \beta)$ and $Q(\gamma, \delta)$ are related only when α or β is equal to one of the stars γ or δ.

In our geometric system G we have taken the points as undefined elements and introduced the lines as certain sets of two or three elements associated with each pair of points. For these lines certain axiomatic properties have been laid down. To complete our geometric system we make the following definition. A subset K of the geometric system G is a *geometric object* when it has the property that for two points P_1 and P_2 in K the collinear point P_3 , when it exists, is also in K.

The common points or cross-cut of two or more geometric objects is again a geometric object because if P_1 and P_2 are points common to them all the collinear point $\beta(P_1 , P_2)$ is also a common point. Furthermore, to any number of geometric objects there exists a minimal geometric object, their union, containing them all. This union may be constructed by successively adjoining all collinear points to the points in the given objects. These remarks show that the geometric objects in G form a complete structure Σ_G . The final step in our theory consists in showing that this geometric structure Σ_G is structurally isomorphic to the structure Σ_S of all partitions of the set S of all stars in G.

We construct this isomorphism by letting correspond to each point P_G in G the minimal partition $P_S = P_S(\alpha_1\beta)$ of S whose only block with more than one element consists of the two stars α and β which P_G connects. Furthermore, to each geometric object A_G we let correspond the union A_S of all minimal partitions P_S which are the images of points P_G contained in A_G . According to the definition of the union of partitions any two stars α and β belong to the same block in A_S if and only if there exists an overlapping chain of blocks with two elements or stars

$$(\alpha, \alpha_1), \qquad (\alpha_1 , \alpha_2), \qquad \cdots , \qquad (\alpha_n , \beta)$$

connecting them. Here any $(\alpha_i , \alpha_{i+1})$ must be a block in a minimal partition $P_S(\alpha_i , \alpha_{i+1})$ which is the image of some point $P_G(\alpha_i , \alpha_{i+1})$ in A_G. But if A_G contains the points $P_G(\alpha_i , \alpha_{i+1})$ and $P_G(\alpha_{i+1} , \alpha_{i+2})$ it follows from the proof of the transitivity of connectedness for stars that A_G also contains the collinear point $P_G(\alpha_i , \alpha_{i+2})$. By induction one concludes therefore that α and β belong to the same block in A_S if and only if the point $P_G(\alpha, \beta)$ belongs to A_G. Conversely, to each partition A_S of S there corresponds a unique subset A_G of G consisting of those points $P_G(\alpha, \beta)$ which correspond to the minimal partitions $P_S(\alpha, \beta)$ contained in A_S. It is clear that A_G must be a geometric object so that we have established a one-to-one correspondence between Σ_G and Σ_S.

It remains to show that this correspondence is a structure isomorphism. We saw that the cross-cut $A_G \cap B_G$ of two geometric objects A_G and B_G consisted of their common points and lines. Thus in the partition of S corresponding to this cross-cut two elements α and β belong to the same block only if they are connected by means of minimal partitions contained both in A_S and B_S. This proves that $A_G \cap B_G$ corresponds to the partition $A_S \cap B_S$.

The union $A_G \cup B_G$ was generated in the following manner. To the set $M_0 = A_G + B_G$ of points in A_G and B_G one adjoins all points $P_3 = \beta(P_1 , P_2)$ which are collinear with a point P_1 in A_G and a point P_2 in B_G. To the larger set M_1 obtained in this way one adjoins further all points $P_3' = \beta(P_1' , P_2')$ which are collinear with a pair of points P_1' and P_2' in M_1. This process may be carried on repeatedly and the union $A_G \cup B_G$ consists of all points which can be constructed by this procedure in a finite number of steps. We now interpret this construction for the partition C_S corresponding to $A_G \cup B_G$ in Σ_S. Let $P_S(\alpha_1 , \alpha_2)$ be the minimal partition corresponding to P_1 and $P_S(\alpha_2 , \alpha_3)$ the partition corresponding to P_2. Then $P_S(\alpha_1 , \alpha_3)$ is the partition corresponding to the collinear point P_3. When this process of construction is continued one sees that any minimal partition $P_S(\alpha, \beta)$ is contained in the partition C_S only if there exists a series of overlapping blocks

$$(\alpha, \alpha_1), \qquad (\alpha_1 , \alpha_2), \qquad \cdots , \qquad (\alpha_n , \beta)$$

connecting α and β. Here each $P_G(\alpha_i, \alpha_{i+1})$ is a point belonging either to A_G or B_G, and hence $P_S(\alpha_i , \alpha_{i+1})$ is a minimal partition belonging either to A_S or B_S. This proves that

$$C_S = A_S \cup B_S$$

in the partition corresponding to $A_G \cup B_G$.

The preceding investigations shall now be summarized as follows.

THEOREM 1. *The geometric system whose lines have the properties postulated in the five enumerated axioms is structurally isomorphic to the structure of all partitions of some set. Conversely the structure of all partitions of a set can be conceived of as being such a geometric system.*

2. **Further remarks upon the characterization of the structure of partitions.** In connection with the preceding characterization of the structure of all partitions we shall make a few further remarks.

First let us mention that the geometrical characterization of the structure of all partitions can also be replaced by criteria which are purely structural in form. Let Σ be a complete structure in which every element is the union of the minimal elements or points which it contains. To establish whether this structure is a structure of all partitions of some set we consider its minimal elements or points. Any pair of points P_1 and P_2 define a line, namely, the set of all points contained in the union $P_1 \cup P_2$. The first two axioms in the preceding theory can then be expressed as follows:

AXIOM 1. *The union $P_1 \cup P_2$ contains at most three points.*

AXIOM 2. *When $P_1 \cup P_2$ contains a third point P_3 then*

$$P_1 \cup P_2 = P_2 \cup P_3 = P_3 \cup P_1 . \tag{3}$$

The two points P_1 and P_2 are related when there exists a third collinear point P_3 such that (3) holds. In this terminology Axioms 3, 4 and 5 of §1 can be translated immediately into structural conditions. We shall not give the explicit reformulation of these axioms. It may also be mentioned that the preceding axiomatic conditions on the points and lines in the structure can be interpreted as graph properties of the diagram representing the structure. The reader may find it interesting to carry through such a theory.

Another procedure consists in expressing the properties of the structure in terms of a multiplicative system. For two points P_1 and P_2 the collinear point $P_3 = P_1 \times P_2$ may be considered to be a product of the two given ones. This product is only defined for certain related pairs of points. The multiplication is found to be commutative and associative and in an equation $P_3 = P_1 \times P_2$ any two elements define the third uniquely. The other axioms are also easily expressible as multiplicative properties.

The dual problem of characterizing the structure of partitions by means of properties of maximal elements is also of interest. We shall not go into details regarding this problem; it may only be mentioned that it can be solved in a fairly simple manner. We have already observed that the structure of all fields of sets over a set S form a structure dually isomorphic to the structure of partitions of S. Consequently the determination of the structure of partitions by means of properties of the maximal elements corresponds to a characterization of the structure of fields of sets over S by means of properties of the minimal fields.

To conclude, let us mention that one can also characterize the structure of equivalence relations by means of properties of arbitrary elements and not only maximal or minimal elements. This can be done, for instance, by means of the singular partitions. These are the distributively indecomposable elements of the given structure and they must correspond in a one-to-one manner to the subsets of the original set S.

3. **Automorphisms.** Let Σ and Σ' denote two structures. A one-to-one correspondence between the elements of these two structures such that

$$a \rightleftarrows a', \qquad b \rightleftarrows b'$$

implies

$$a \cap b \rightleftarrows a' \cap b', \qquad a \cup b \rightleftarrows a' \cup b'$$

is called an *isomorphism*. An isomorphism of a structure to itself is called an *automorphism*. We shall now determine the automorphisms of the structure of equivalence relations on a set S.

The solution of this problem follows from a simple study of the properties of maximal partitions. A maximal partition $P_A = P_{\overline{A}}$ of a set S is a partition with two blocks A and $\overline{A} = S - A$. The cross-cut $P_A \cap P_B$ of two maximal partitions P_A and P_B is the partition with the four blocks

$$A \cdot B, \qquad A \cdot \overline{B}, \qquad \overline{A} \cdot B, \qquad \overline{A} \cdot \overline{B}. \tag{4}$$

As before let us denote the universal partition of S by U. If none of the blocks in (4) are void it is seen that the quotient structure

$$U/P_A \cap P_B \tag{5}$$

is isomorphic to the structure of all partitions of a set with four elements. Thus the structure (5) contains 15 elements including $P_A \cap P_B$ and U.

There can be a void set in (4) only when $A \supset B$ or $\overline{A} \supset B$. In this case we shall say that the maximal partitions P_A and P_B are *related*. When P_A and P_B are related the quotient structure (5) is isomorphic to the structure of partitions of a set with three elements.

The special maximal partitions P_a in which the block $A = a$ is a single element are characterized by the property that they are related to all other maximal partitions. It is therefore clear that by any automorphism the singular maximal partition P_a must be transformed into some other singular maximal partition $P_{a'}$. We have seen that every partition is the direct union of singular partitions. Furthermore, every singular partition can be represented as the intersection of singular maximal partitions. From these remarks we deduce that every automorphism which leaves all singular maximal partitions invariant is the unit automorphism. Thus every automorphism can be represented uniquely as a permutation of the partitions P_a, hence as a permutation of the set S. But, conversely, every permutation of the set S must correspond to an automorphism of the structure of partitions of S so that we have the following result, due to Garrett Birkhoff [1; Theorem 23] for the case of a finite set.

THEOREM 2. *The group of automorphisms of the structure of equivalence relations on a set S is isomorphic to the symmetric group defined by the elements in S.*

It may be mentioned in this connection that the preceding characterization of the singular maximal partitions may be used to obtain an axiomatic characterization of the structure of partitions by means of properties of maximal partitions.

4. **Homomorphisms in structures.** A structure Σ is said to be *homomorphic* to a structure Σ^α if there exists a correspondence α from Σ to Σ^α such that every element in Σ^α is the image of at least one element in Σ and furthermore such that

$$a \to a^\alpha, \qquad b \to b^\alpha$$

implies

$$a \cup b \to a^\alpha \cup b^\alpha, \qquad a \cap b \to a^\alpha \cap b^\alpha.$$

The correspondence α itself is called a *homomorphism* from Σ to Σ^α. When Σ^α is a substructure of Σ we call α an *endomorphism* of Σ.

In the following, we are particularly interested in the homomorphisms of the structure of all partitions of a set, but in order to solve the problem of finding all homomorphisms of this special structure it is necessary to make some general observations on homomorphisms for arbitrary structures.

We observe first that by a homomorphism α of a structure Σ to a structure Σ^α those elements a which correspond to the same image a^α in Σ^α form a substructure Σ_{a^α} of Σ. If, namely,

$$a_1 \to a^\alpha, \qquad a_2 \to a^\alpha,$$

then

$$a_1 \cup a_2 \to a^\alpha, \qquad a_1 \cap a_2 \to a^\alpha.$$

Furthermore, one sees that Σ_{a^α} is a *dense* substructure of Σ, i.e., if Σ_{a^α} contains two elements $a_1 \supset a_2$ it contains every element between a_1 and a_2. Thus every homomorphism of a structure corresponds to a partition of its elements into dense substructures. It should be observed, however, that not every such partition need define a homomorphism. When the finite chain condition is satisfied in Σ the substructures Σ_{a^α} are quotient structures. The unit element e of Σ corresponds to the unit element e^α of Σ^α but Σ^α can have a unit element even when Σ does not.

We shall now turn to a particular construction for homomorphisms. Let Σ_0 be a substructure of Σ. We shall write

$$a_1 \equiv a_2 \pmod{\Sigma_0}$$

and say that a_1 and a_2 are *equivalent with respect to* Σ_0 if there exist elements t_1 and t_2 in Σ_0 such that

$$a_1 \cup t_1 = a_2 \cup t_2 .$$

It is obviously no restriction to assume that Σ_0 contains all elements s in Σ for which there exists an element t_0 in Σ_0 such that $t_0 \supset s$. The condition for equivalence with respect to Σ_0 may also be written

$$a_1 \cup t_0 = a_2 \cup t_0 ,$$

where t_0 shall belong to Σ_0 . One sees immediately that the equivalence thus defined satisfies the three axioms for an equivalence relation. From the cross-division of Σ (mod Σ_0) one obtains a correspondence α by letting each element a correspond to the class $\{a\}$ to which it belongs

$$a \to a^\alpha = \{a\}.$$

One observes that

$$a_1 \equiv a_2 , b_1 \equiv b_2 \pmod{\Sigma_0} \tag{6}$$

implies

$$a_1 \cup b_1 \equiv a_2 \cup b_2 \pmod{\Sigma_0}.$$

This shows that one can define the union of two classes by putting

$$\{a\} \cup \{b\} = \{a \cup b\}$$

and the union is independent of the choice of the representatives a and b in their respective classes.

For the cross-cut of two classes a similar definition is usually not possible. One sees, however, that if the congruences (6) always imply

$$a_1 \cap b_1 \equiv a_2 \cap b_2 \pmod{\Sigma_0} \tag{7}$$

then one can define uniquely

$$\{a\} \cap \{b\} = \{a \cap b\}$$

and in this case the congruence classes (mod Σ_0) form a structure which is homomorphic to Σ. Such a homomorphism shall be called a *modular homomorphism*.

We shall now determine the conditions for a substructure Σ_0 to define a modular homomorphism. Since one always has

$$a \equiv a \cup t \pmod{\Sigma_0}$$

when t belongs to Σ_0 it follows that one must always have

$$a \cap b \equiv (a \cup t) \cap b \pmod{\Sigma_0} \tag{8}$$

for an arbitrary pair of elements a and b. But one can show conversely that when the condition (8) is always satisfied the congruences (6) must imply that (7) holds. One has, namely, according to (6),

$$a_1 \cup t_1 = a_2 \cup t_1 , \qquad b_1 \cup t_2 = b_2 \cup t_2 ,$$

where t_1 and t_2 belong to Σ_0. Through the application of (8) one obtains successively

$$a_1 \cap b_1 \equiv (a_1 \cup t_1) \cap b_1 \equiv (a_2 \cup t_1) \cap b_1 \equiv a_2 \cap b_1 \pmod{\Sigma_0}$$

and similarly

$$a_1 \cap b_1 \equiv a_2 \cap b_1 \equiv a_2 \cap b_2 \pmod{\Sigma_0}.$$

We can state, therefore,

THEOREM 3. *The necessary and sufficient condition for a substructure Σ_0 of Σ to define a modular homomorphism is that for every pair of elements a and b in Σ and for every element t in Σ_0 there exist an element t' in Σ_0 such that*

$$(9) \qquad (a \cap b) \cup t' = ((a \cup t) \cap b) \cup t'.$$

It should be observed that t' will usually depend on all three of the elements a, b and t. It is always permissible to assume in (9) that the element t' contains t since otherwise one could replace t' by $t' \cup t$.

Condition (9) of Theorem 3 can also be stated as follows. For every pair of elements a and b in Σ and t in Σ_0 there will exist some $t' \supset t$ in Σ_0 such that

$$(10) \qquad ((a \cup t) \cap (b \cup t)) \cup t' = (a \cap b) \cup t'.$$

To prove that condition (10) implies (9) we need only to observe that

$$((a \cup t) \cap (b \cup t)) \cup t' \supset ((a \cup t) \cap b) \cup t' \supset (a \cap b) \cup t'.$$

To show conversely that condition (10) implies (9) we write

$$((a \cup t) \cap (b \cup t)) \cup t_1 = (a \cap (b \cup t)) \cup t_1,$$

$$(a \cap (b \cup t)) \cup t_2 = (a \cap b) \cup t_2$$

and find that (10) holds with $t' = t_1 \cup t_2$.

Let us consider the special case where Σ_0 has a universal element m_0. Such an element always exists when the ascending chain condition holds in Σ. We have then

$$a \equiv b \pmod{\Sigma_0}$$

if and only if

$$a \cup m_0 = b \cup m_0.$$

In this case we can also write

$$a \equiv b \pmod{m_0}$$

and Σ_0 is the structure of elements in Σ contained in m_0. The condition for an element m_0 to define a modular homomorphism may then be stated simply that for every pair of elements a and b in Σ one will have

$$(a \cup m_0) \cap (b \cup m_0) = (a \cap b) \cup m_0 \tag{11}$$

or also

$$((a \cup m_0) \cap b) \cup m_0 = (a \cap b) \cup m_0 . \tag{12}$$

An element m_0 such that relations (11) or (12) hold for all elements a and b in Σ will be called *distributive*. Any distributive element defines a modular homomorphism

$$a \to a \cup m_0$$

which projects Σ upon the structure of all elements in Σ containing m_0.

In order to determine all modular homomorphisms of a given structure Σ it is necessary to find all substructures Σ_0 which have the property required by Theorem 3. We observed that one can always assume that Σ_0 contains every element x for which $t \supset x$ for some element t in Σ_0. The substructure Σ_0 is therefore completely determined when one has found a set $\mathfrak{S}(t_i)$ of elements t_i in Σ_0 such that every other element t in Σ_0 is contained in at least one of the elements t_i. Such a set $\mathfrak{S}(t_i)$ in Σ_0 may be called an *enveloping set*. One sees that in order to verify (9) for a substructure Σ_0 it is sufficient to verify it for elements t and t' belonging to an enveloping set.

The preceding theory may de dualized. We let Σ_0 denote a substructure of Σ with the property that when t is in Σ_0 every element $x \supset t$ is also in Σ_0. We write

$$a_1 \equiv a_2 \pmod{\Sigma_0}$$

when

$$a_1 \cap t = a_2 \cap t$$

for some element t in Σ_0. This definition of congruences introduces a class-division of Σ and as before these classes constitute a system Σ^a which under certain conditions is a structure homomorphic to Σ. This type of homomorphism shall be called a *dual modular homomorphism*. As in Theorem 3 one finds that Σ_0 defines a dual modular homomorphism if and only if for any pair of elements a and b in Σ and any element t in Σ_0 one can find an element $t' \subset t$ in Σ_0 such that

$$(a \cup b) \cap t' = ((a \cap t) \cup b) \cap t'. \tag{13}$$

When Σ_0 has a unit element n_0 the necessary and sufficient condition for a dual modular homomorphism becomes

$$(a \cap n_0) \cup (b \cap n_0) = (a \cup b) \cap n_0 . \tag{14}$$

An element n_0 for which this relation (14) holds for all a and b in Σ shall be called a *dual distributive element*. Any dual distributive element defines a dual modular homomorphism through the correspondence

$$a \to a \cap n_0 .$$

5. **Homomorphisms of the structure of partitions.** The preceding general theory will now be used to determine the homomorphisms of the structure Σ of all partitions of a set S. As a simple consequence of Theorems 13 and 15 of Chapter 1, we have

THEOREM 4. *Except for the unit partition E and the universal partition U there exists no distributive or dual distributive element in the structure of partitions.*

This theorem shows that in the case of a finite set S no modular or dual modular homomorphism except the trivial ones can exist. We can therefore exclude the finite case from the subsequent considerations if we so desire. The study of the infinite case is based upon the following

LEMMA. *Any substructure Σ_0 defining a modular homomorphism in the structure Σ of all partitions has an enveloping set consisting of singular partitions.*

Proof. We shall have to show that every partition in Σ_0 is contained in a singular partition also belonging to Σ_0. Let T be some non-singular partition in Σ_0. Those blocks in T which contain more than one element shall be denoted by T_i and their set sum by

$$T_0 = \sum_i T_i .$$

In each block T_i we select two different elements a_i and a_i' and we denote the sets of all a_i and a_i' by A_0 and A_0' respectively. Furthermore, A and A' are the singular partitions whose main blocks are A_0 and A_0'. Finally, B and B' are the singular partitions with the main blocks

$$B_0 = T_0 - A_0 , \qquad B_0' = T_0 - A_0' .$$

According to these definitions one finds that

$$T^* = A \cup T = A' \cup T$$

is the singular partition whose main block is T_0. We shall prove our lemma by showing that T^* belongs to Σ_0. According to (9) there must exist a partition T' in Σ_0 such that

$$(A \cap B) \cup T' = ((A \cup T) \cap B) \cup T'.$$

But in this case one finds

$$(A \cup T) \cap B = B, \qquad A \cap B = E$$

so that there must exist some T' in Σ_0 such that $T' \supset B$. Similarly one deduces the existence of a partition T'' in Σ_0 such that $T'' \supset B'$. Therefore, B and B' both belong to Σ_0 and since

$$T^* = T \cup B \cup B'$$

our lemma is proved.

From this lemma we deduce

THEOREM 5. *The structure Σ of all partitions of a set S has no modular homomorphisms except the trivial ones defined by the unit partition E and the universal partition U.*

Proof. It is sufficient to show that if the substructure Σ_0 which defines a modular homomorphism contains a partition different from E then U belongs to Σ_0 so that $\Sigma_0 = \Sigma$. It is clear that if Σ_0 contains any partition different from E it must contain some minimal singular partition P whose main block P_0 consists of two elements a_1 and a_2. We have already observed that the basic set S can be assumed to have an infinite cardinal number σ. Then it is possible to define two disjoint sets A_0 and A_0' such that

$$S = A_0 + A_0'$$

and a_1 belongs to A_0 and a_2 to A_0'. The maximal partition whose blocks are A_0 and A_0' shall be denoted by A so that

$$A \cup P = U.$$

Since A_0 and A_0' have the same cardinal number, a one-to-one correspondence

$$a \rightleftarrows a'$$

can be established between the elements of the two sets. This correspondence can be used to define a partition B whose blocks are pairs of elements (a, a'). One sees that

$$(A \cup P) \cap B = B, \qquad A \cap B = E,$$

and when (9) is applied it follows that there must exist a partition P' in Σ_0 containing B. Consequently, B belongs to Σ_0 and according to the preceding lemma there must exist a singular partition in Σ_0 containing B. But since B has no blocks consisting of single elements the only singular partition which contains B is U and Theorem 5 is proved.

One can also deduce the dual theorem:

THEOREM 6. *The structure Σ of all partitions has no dual modular homomorphisms except the trivial ones.*

This result may be proved as a consequence of the relation (13) which gives the condition for a dual modular homomorphism, but the proof will be omitted since the theorem is also a consequence of the next theorem.

In order to complete these investigations on the homomorphisms of the structure of partitions we shall establish a rather interesting theorem on homomorphisms of structures in which there exist relative or dual relative complements. Let us recall that a structure Σ has relative complements if to every pair $a \supset b$ of elements in Σ there exists an element c such that

$$a \cup c = u, \qquad a \cap c = b$$

and Σ has dual relative complements if there exists a d such that

$$b \cup d = a, \qquad b \cap d = e.$$

Here u and e are the universal and unit element as before. We prove

THEOREM 7. *In a structure with dual relative complements every homomorphism is a modular homomorphism.*

Proof. Let α be a homomorphism of Σ to Σ^α and let Σ_{e^α} be the structure of all elements in Σ having the same image in Σ^α as the unit element e. Next let a and b be two elements with the same image a^α in Σ^α. Since $a \cup b$ also has the image a^α it is no limitation to assume that $a \supset b$. Because there exist dual relative complements some d can be determined such that

$$b \cup d = a, \qquad b \cap d = e.$$

When α is applied to these two relations one obtains

$$a^\alpha \cup d^\alpha = a^\alpha, \qquad a^\alpha \cap d^\alpha = e^\alpha.$$

The first of these conditions states that d^α is contained in a^α and the second gives $d^\alpha = e^\alpha$. Thus $a^\alpha = b^\alpha$ always implies

$$a \equiv b \pmod{\Sigma_{e^\alpha}}$$

and conversely; hence α is a modular homomorphism.

The dual Theorem 7 is obtained analogously. Theorem 6 is seen to be a consequence of Theorems 5 and 7. The combination of these two theorems also leads to the principal result, which is

THEOREM 8. *The structure of equivalence relations has no homomorphisms except isomorphisms and the trivial homomorphism in which every element has the same image.*

An *endomorphism* of a structure Σ is a homomorphism of Σ to a substructure Σ^α. Since the structure of partitions has no essential homomorphisms it cannot have any endomorphisms when the basic set S is finite. But when S is infinite there always exist certain endomorphisms of Σ. These endomorphisms may be

determined by a method based upon the classification of the structure of partitions given previously. Since the analysis involved is rather cumbersome it shall be left to an interested reader.

Bibliography

1. Garrett Birkhoff, *On the structure of abstract algebras*, Proceedings of the Cambridge Philosophical Society, vol. 31(1935), pp. 433–454.
2. Paul Dubreil and Marie-Louise Dubreil-Jacotin, *Propriétés algébriques des relations d'équivalence*, Comptes Rendus, Paris, vol. 205(1937), pp. 704–706.
3. Paul Dubreil and Marie-Louise Dubreil-Jacotin, *Propriétés algébriques des relations d'équivalence; théorèmes de Schreier et de Jordan-Hölder*, Comptes Rendus, Paris, vol. 205(1937), pp. 1349–1351.
4. Paul Dubreil and Marie-Louise Dubreil-Jacotin, *Théorie algébrique des relations d'équivalence*, Journal de Mathématiques, (9), vol. 18(1939), pp. 63–95.
5. L. F. Epstein, *A function related to the series for* exp (exp x), Journal of Mathematics and Physics, vol. 18(1939), pp. 153–173.
6. F. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig, 1914.
7. Dénes König, *Ueber Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre*, Mathematische Annalen, vol. 77(1916), pp. 453–465.
8. W. Maak, *Eine neue Definition der fastperiodischen Funktionen*, Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität, vol. 11(1936), pp. 240–244.
9. G. A. Miller, *On a method due to Galois*, Quarterly Journal of Mathematics, vol. 41(1910), pp. 382–384.
10. G. A. Miller, *Same left co-set and right co-set multipliers for any given finite group*, Bulletin of the American Mathematical Society, vol. 29(1923), pp. 394–398.
11. Oystein Ore, *Remarks on structures and group relations*, Naturf. Gesellschaft Zürich, vol. 85(1940), Festschrift Rudolf Fueter, pp. 1–4.
12. Oystein Ore, *On the theorem of Jordan-Hölder*, Transactions of the American Mathematical Society, vol. 41(1937), pp. 266–275.
13. Oystein Ore, *Theory of monomial groups*, Transactions of the American Mathematical Society, vol. 51(1942), pp. 15–64.
14. V. Schmushkovitch, *On a combinatorial theorem of the theory of sets*, in Russian, Recueil Mathématique, Moscou (Mat. Sbornik) N. S. 6 (48), (1939), pp. 139–147.
15. Shien-siu Shü, *On the common representative system of residue classes of infinite groups*, Journal of the London Mathematical Society, vol. 16(1941), pp. 101–104.
16. W. Sierpiński, *Sur le plus petit corps contenant une famille donnée d'ensembles*, Fundamenta Mathematicae, vol. 30(1938), pp. 14–16.

Yale University.

THE RECIPROCAL OF CERTAIN TYPES OF HURWITZ SERIES

By L. Carlitz

1. **Introduction.** By (integral) Hurwitz series we shall mean series of the form

$$f(u) = \sum_{m=1}^{\infty} \frac{A_m}{g_m} u^m, \tag{1.1}$$

where $A_m = A_m(x)$ is a polynomial in an indeterminate x with coefficients in a fixed Galois field $GF(p^n)$, and the denominator is defined by

$$g_m = g(m) = F_1^{m_1} F_2^{m_2} \cdots F_s^{m_s} \qquad (g_0 = 1),$$

where

$$m = a_0 + a_1 p^n + \cdots + a_s p^{ns} \qquad (0 \leq a_i < p^n),$$

and

$$F_k = (x^{p^{nk}} - x)(x^{p^{nk}} - x^{p^n}) \cdots (x^{p^{nk}} - x^{p^{n(k-1)}}) \qquad (F_0 = 1).$$

For properties of Hurwitz series required here see [1; esp. 507-509]. We are interested in the coefficients of $u/f(u)$.

First, however, we consider the "linear" case

$$f(u) = \sum_{i=0}^{\infty} \frac{A_i}{F_i} u^{p^{ni}} \qquad (A_0 = 1), \tag{1.2}$$

where again A_i is integral, that is, a polynomial in x. Now the inverse of (1.2) has the same general form, say

$$\lambda(u) = \sum_{i=0}^{\infty} \frac{E_i}{F_i} u^{p^{ni}} \qquad (E_0 = 1), \tag{1.3}$$

with integral E_i. In a previous paper [2] we assumed

$$E_i \equiv 0 \pmod{g(p^{ni} - 1)}. \tag{1.4}$$

We now make the milder assumption

$$E_i \equiv 0 \pmod{L_{i-1}} \tag{1.5}$$

for $i > 0$, where

$$L_k = (x^{p^{nk}} - x)(x^{p^{n(k-1)}} - x) \cdots (x^{p^n} - x) \qquad (L_0 = 1).$$

Define β_m by means of

$$\frac{u}{f(u)} = \sum_{m=0}^{\infty} \beta_m \frac{u^m}{g_m}$$

Received April 9, 1942.

so that β_m is rational in x. Then we prove that

$$\beta_m = G_m - eE_h' \sum_k (-1)^{k-1} E_k^d L_{k-1} \sum_P \frac{1}{P}, \tag{1.6}$$

where G_m is integral, $E_h = E_h' L_{h-1}$, e and d are rational integers ($p \nmid e$), the inner summation is over irreducible polynomials P of degree k, and h and k are integers determined by m and satisfying certain conditions. If these conditions are not satisfied, then (1.6) reduces to $\beta_m = G_m$, that is, β_m is integral. It will be noted that (1.6) is somewhat more complicated than the corresponding theorem based on (1.4); in the present case, however, it is not necessary to frame a supplementary theorem for $p^n = 2$.

Turning next to the general series (1.1) we denote the inverse function by

$$\lambda(u) = \sum_{m=1}^{\infty} \epsilon_m \frac{u^m}{g_m} \qquad (A_1 = \epsilon_1 = 1).$$

Corresponding to (1.5) we now assume

$$L_i \mid \epsilon_m \qquad \text{for } p^{ni} < m. \tag{1.7}$$

The final main result is similar to (1.6). The following formula enables us to make use of the results of the linear case:

$$\frac{1}{F_k} f^{p^{nk}}(u) \equiv \sum_{i=0}^{\infty} A_{p^{nki}} \frac{u^{p^{nk(i+1)}}}{F_{k(i+1)}} \pmod{P}, \tag{1.8}$$

where P is irreducible of degree k. By $\sum_m A_m \frac{u^m}{g_m} \equiv \sum_m A_m' \frac{u^m}{g_m} \pmod{P}$ is meant the system of congruences $A_m \equiv A_m'$ ($m = 1, 2, \cdots$). We remark that (1.8) is independent of the hypothesis (1.7).

2. **Preliminary results for the linear case.** Substituting from (1.2) and (1.3) in the identity

$$\lambda(f(u)) = u, \tag{2.1}$$

we have at once

$$\sum_{i=0}^{m} \begin{Bmatrix} m \\ i \end{Bmatrix} E_i A_{m-i}^{p^{ni}} = 0 \qquad (m > 0), \tag{2.2}$$

where the coefficient $\begin{Bmatrix} m \\ i \end{Bmatrix}$ is defined by

$$\begin{Bmatrix} m \\ i \end{Bmatrix} = \frac{F_m}{F_i F_{m-i}^{p^{ni}}}, \quad \begin{Bmatrix} m \\ 0 \end{Bmatrix} = \begin{Bmatrix} m \\ m \end{Bmatrix} = 1, \tag{2.3}$$

and is integral. Now let P denote an irreducible polynomial of degree k. For the present purpose it will be sufficient to assume

$$P \mid E_i \qquad \text{for } i > k, \tag{2.4}$$

which is weaker than (1.5). Then using (2.4), (2.2) becomes

$$\sum_{i=0}^{k} \begin{Bmatrix} m \\ i \end{Bmatrix} E_i \, A_{m-i}^{p^{ni}} \equiv 0 \pmod{P}. \tag{2.5}$$

Now by (2.3)

$$\begin{Bmatrix} m \\ i \end{Bmatrix} = \frac{[m]\,[m-1]^{p^n} \cdots [m-i+1]^{p^{n(i-1)}}}{F_i} \qquad ([m] = x^{p^{nm}} - x), \tag{2.6}$$

so that, for $k \mid m$,

$$\begin{Bmatrix} m \\ i \end{Bmatrix} \equiv \begin{cases} 0 \text{ for } 0 < i < k, \\ 1 \text{ for } i = 0, k. \end{cases}$$

Hence, if we replace m by km, (2.5) reduces to

$$A_{km} + E_k A_{k(m-1)}^{p^{nk}} \equiv 0,$$

that is,

$$A_{km} \equiv -E_k A_{k(m-1)}, \qquad A_k \equiv -E_k.$$

Repeated application of this recursion leads to

$$A_{km} \equiv (-1)^m E_k^m \equiv A_k^m \pmod{P}. \tag{2.7}$$

Next consider

$$\frac{f^{p^{nk}}}{F_k} = \sum_{m=k}^{\infty} \begin{Bmatrix} m \\ k \end{Bmatrix} A_{m-k}^{p^{nk}} \frac{u^{p^{nm}}}{F_m}.$$

Using (2.6) it is easily seen that

$$\begin{Bmatrix} m \\ k \end{Bmatrix} \equiv \begin{cases} 0 \text{ for } k \nmid m, \\ 1 \text{ for } k \mid m; \end{cases}$$

then by (2.7) we get

$$\frac{f^{p^{nk}}}{F_k} \equiv \sum_{i=1}^{\infty} A_k^{i-1} \frac{u^{p^{nki}}}{F_{ki}}. \tag{2.8}$$

Again from (2.1) and (2.4) it follows that

$$u \equiv \sum_{i=0}^{k} \frac{E_i}{F_i} f^{p^{ni}} \pmod{P}. \tag{2.9}$$

For brevity, put

$$f_0 = u - \frac{E_k}{F_k} f^{p^{nk}}; \tag{2.10}$$

then by (2.8)

$$f_0 \equiv u + \frac{A_k}{F_k} f^{p^{nk}} \equiv \sum_{i=0}^{\infty} A_k^i \frac{u^{p^{nki}}}{F_{ki}}. \tag{2.11}$$

Next, from (2.9) and (2.10) it follows that

$$f_0 \equiv f + \sum_{i=1}^{k-1} \frac{E_i}{F_i} f^{p^{ni}}.$$

Raising both members of this congruence to the $(p^{nk}-1)$-th power we get

$$f_0^{p^{nk}-1} \equiv f^{p^{nk}-1} + R,$$

where R denotes a sum of terms each of which involves f^r, where $r \geq p^{nk}$. Hence [1; Theorem 3] $R \equiv 0$ and we have the following [cf. 2; 237, formula (2.13)]

LEMMA 1. *If the coefficients of* (1.3) *satisfy* (2.4), *then*

$$f^{p^{nk}-1} \equiv \left(\sum_{i=0}^{\infty} A_{ki} \frac{u^{p^{nki}}}{F_{ki}}\right)^{p^{nk}-1} \pmod{P}, \tag{2.12}$$

where P is irreducible of degree k.

When (1.4) holds, this result suffices for the main theorem. For the weaker assumption (1.5), however, it is necessary to extend (2.12). We shall require a formula for

$$\frac{f^{p^{nks}-1}}{g(p^{nks}-1)} \pmod{P}. \tag{2.13}$$

In the first place, by (2.10)

$$E_k \frac{f^{p^{nk}}}{F_k} \equiv f_0 - u. \tag{2.14}$$

Now if $f_1(u) \equiv f_2(u)$, where f_1 and f_2 are arbitrary series of the form (1.1), it is easy to see that

$$\frac{f_1^{p^{nk}}}{F_k} \equiv \frac{f_2^{p^{nk}}}{F_k} \pmod{P}.$$

Hence, (2.14) yields

$$E_k^2 \frac{f^{p^{2nk}}}{F_{2k}} \equiv \frac{E_k}{F_k}\left(E_k \frac{f^{p^{nk}}}{F_k}\right)^{p^{nk}} \equiv \frac{E_k}{F_k}(f_0 - u)^{p^{nk}} \equiv f_0 - u - \frac{E_k}{F_k} u^{p^{nk}}.$$

Continuing in this way we get

$$E_k^i \frac{f^{p^{nki}}}{F_{ki}} \equiv f_0 - u - \cdots - E_k^{i-1} \frac{u^{p^{nk(i-1)}}}{F_{k(i-1)}}. \tag{2.15}$$

From (2.12) and (2.15) follows

$$E_k^{p^{nks}-1}\frac{f^{p^{nks}-1}}{(F_k\cdots F_{k(s-1)})^{p^{nk}-1}}$$

$$\equiv f_0^{p^{nk}-1}(f_0-u)^{p^{nk}-1}\left(f_0-u-\frac{E_k}{F_k}u^{p^{nk}}\right)^{p^{nk}-1}\cdots.$$

If $P \nmid E_k$, this reduces to

$$\frac{f^{p^{nks}-1}}{g(p^{nks}-1)} \equiv \frac{f_0^{p^{nk}-1}}{g(p^{nk}-1)}\frac{u^{p^{nk(s-1)}-1}}{g(p^{nk(s-1)}-1)}; \tag{2.16}$$

it is easily seen that (2.16) holds for $P \mid E_k$ also. Put

$$\frac{f_0^{p^{nk}-1}}{g(p^{nk}-1)} = \sum_m C_m\frac{u^m}{g_m}, \tag{2.17}$$

so that the right member of (2.16) becomes

$$\sum_m C_m\frac{g(m+p^{nkt}-1)}{g(m)g(p^{nkt}-1)}\frac{u^{m+p^{nkt}-1}}{g(m+p^{nkt}-1)} \qquad (t=s-1).$$

Now it may be verified that

$$\frac{g(m+p^{nkt}-1)}{g(m)g(p^{nkt}-1)} \equiv \begin{cases}1 \text{ for } p^{nkt}\mid m,\\ 0 \text{ otherwise.}\end{cases} \tag{2.18}$$

Clearly, then, the only terms in the right member of (2.17) that we need consider are those arising in the expansion of

$$\left(\sum_{i=t}^{\infty}A_{ki}\frac{u^{p^{nki}}}{F_{ki}}\right)^{p^{nk}-1}.$$

We may now state

LEMMA 2. *If $f(u)$ satisfies the hypothesis of Lemma* 1, *then*

$$\frac{f^{p^{nks}-1}}{g(p^{nks}-1)} \equiv \frac{1}{g(p^{nk}-1)}\left(\sum_{i=s-1}^{\infty}A_{ki}\frac{u^{p^{nki}}}{F_{ki}}\right)^{p^{nk}-1}\frac{u^{p^{nk(s-1)}-1}}{g(p^{nk(s-1)}-1)}. \tag{2.19}$$

It is now not difficult to expand (2.13). Define $\mu(m)$ by means of

$$\begin{aligned} m &= b_0+b_1p+\cdots+b_rp^r \qquad (0\le b_i<p),\\ \mu(m) &= b_0+b_1+\cdots+b_r. \end{aligned} \tag{2.20}$$

Then it may be shown that [2; end of §2] for C_m as defined by (2.17) we have $C_m \equiv 0$ unless

$$p^{nk}-1 \mid m, \qquad \mu(m)=nk(p-1), \tag{2.21}$$

while if both parts of (2.21) are satisfied then

$$C_m \equiv \frac{(-1)^{nk}}{\prod b_i!} A_k^d \pmod{P}, \tag{2.22}$$

where

$$d = \sum_{i,j} i p^j b_{nki+j}.$$

To determine the non-vanishing terms in the right member of (2.19), consider the term in u^m. Write

$$m = p^{nk(s-1)} - 1 + m_0. \tag{2.23}$$

Also (2.21) must hold for m_0 :

$$p^{nk} - 1 \mid m_0, \qquad \mu(m_0) = nk(p-1).$$

Hence it follows that

$$\begin{aligned}\mu(m) &= \mu(p^{nk(s-1)} - 1) + \mu(m_0) \\ &= nk(s-1)(p-1) + nk(p-1) = nks(p-1).\end{aligned}$$

Conversely, suppose $\mu(m) = nh(p-1)$; then (2.19) will contribute provided

$$h = ks, \qquad p^{nk} - 1 \mid m, \qquad p^{nk(s-1)} \mid m + 1 \tag{2.24}$$

hold. If (2.24) is satisfied, then the contribution is C_{m_0}. This proves

LEMMA 3. *If* (2.4) *holds and* $k \mid h$, *then*

$$\frac{f^{p^{nk}-1}}{g(p^{nk}-1)} \equiv \sum_m C_{m_0} \frac{u^m}{g_m} \pmod{P}, \tag{2.25}$$

where C_{m_0} *is defined by* (2.22) *and* (2.23).

3. **The main theorem for the linear case.** Clearly

$$\frac{u}{f} = \frac{\lambda(f)}{f} = \sum_{h=0}^{\infty} \frac{E_h}{F_h} f^{p^{nh}-1} = \sum_{h=0}^{\infty} \frac{E_h}{L_h} \frac{f^{p^{nh}-1}}{g(p^{nh}-1)}. \tag{3.1}$$

If we put

$$\frac{f^{p^{nh}-1}}{g(p^{nh}-1)} = \sum_m C_m^{(h)} \frac{u^m}{g_m},$$

where $C_m^{(h)}$ is integral, then comparison with (3.1) gives

$$\beta_m = \sum_h \frac{E_h}{L_h} C_m^{(h)}, \tag{3.2}$$

the summation extending over all h such that $p^{nh} - 1 \leq m$.

Now assume (1.5), that is, put

$$E_h = E'_h L_{h-1} \qquad (h \geq 1), \qquad E'_0 = 1, \tag{3.3}$$

so that (3.2) becomes

$$\beta_m = \sum_h \frac{E'_h}{[h]} C_m^{(h)} \qquad ([h] = x^{p^{nh}} - x), \tag{3.4}$$

from which it is clear that the denominator of β_m contains only simple factors (at most). For more precise information we make use of the results of §2. Let m be fixed and P irreducible of degree k. Then P can occur in the denominator of β_m only if $\mu(m) = nh(p-1)$ and in addition the several parts of (2.24) are all satisfied. Since h is uniquely determined by m it follows that (3.4) reduces to

$$\beta_m = G_m + \frac{E'_h}{[h]} C_m^{(h)} \qquad (\mu(m) = nh(p-1)), \tag{3.5}$$

where G_m is integral. If $\mu(m)$ is not divisible by $n(p-1)$, then (3.5) becomes simply $\beta_m = G_m$, that is, β_m is integral. If, however, $\mu(m) = nh(p-1)$, then according to (3.5) the denominator of β_m may contain irreducibles P of degree k, where $h = ks$. However, we have seen that $C_m^{(h)} \equiv 0 \pmod{P}$ unless

$$p^{nh} - 1 \mid m, \qquad p^{nh(s-1)} \mid m + 1. \tag{3.6}$$

Assume that (3.6) is satisfied; then since

$$-\frac{1}{[h]} = \sum_{\deg P \mid h} \frac{P'}{P},$$

the summation extending over all P of degree k dividing h, we get

$$-\frac{C_m^{(h)}}{[h]} = G_m + C_m^{(h)} \sum_P \frac{P'}{P}, \tag{3.7}$$

where the summation is now restricted to k satisfying (3.6). Substituting from (3.7) in (3.5) we have finally

$$\beta_m = G_m - E'_h C_m^{(h)} \sum_P \frac{P'}{P}.$$

Now, use (2.22) and (2.7) and we obtain the following

THEOREM 1. *Assume the series*

$$f(u) = \sum_{i=0}^{\infty} \frac{A_i}{F_i} u^{p^{ni}} \qquad (A_0 = 1)$$

has an inverse satisfying (1.5). *Put* $E_h = E'_h L_{h-1}$,

$$\frac{u}{f(u)} = \sum_{m=0}^{\infty} \beta_m \frac{u^m}{g_m} \qquad (\beta_0 = 1).$$

If the system

$$\mu(m) = nh(p-1), \qquad h = ks, \qquad p^{nk}-1 \mid m, \qquad p^{nk(s-1)} \mid m+1 \tag{3.8}$$

(where $\mu(m)$ is defined as in (2.20)) is inconsistent, then β_m is integral, while if (3.8) is consistent then

$$\beta_m = G_m - eE_k' \sum_{k \mid h} E_k^d \sum_{\deg P = k} \frac{P'}{P}, \tag{3.9}$$

where G_m is integral, the inner summation is over all irreducible P of degree k and the outer summation is over all k satisfying (3.8), and d and e are determined by

$$m - p^{nk(s-1)} + 1 = p^{nk(s-1)} \sum_i b_i p^i \qquad (0 \le b_i < p),$$

$$e = \frac{(-1)^{nk+d}}{\prod b_i!}, \qquad d = \sum_{i,j} i p^j b_{nki+j}\,.$$

The outer summation will extend over all k dividing h only when $m = p^{nh} - 1$. In this case there is the explicit formula $\beta_{p^{nh}-1} = E_h/L_h$.

Since $P' \equiv (-1)^{k-1} L_{k-1} \pmod P$, we have as a variant of (3.9):

$$\beta_m = G_m - eE_k' \sum_k (-1)^{k-1} E_k^d L_{k-1} \sum_P \frac{1}{P}, \tag{3.10}$$

which is the result stated in the introduction.

One or two immediate corollaries of Theorem 1 may be mentioned. In the first place, it follows at once that

$$\beta_{m,x} = x(x^m - 1)\beta_m$$

is integral; more generally

$$\beta_{m,U} = U(U^m - 1)\beta_m$$

is integral for U an arbitrary polynomial. Second, if $\mu(m) = nh(p-1)$, then

$$[h]\beta_m = (x^{p^{nh}} - x)\beta_m$$

is integral, and generally

$$(U^{p^{nh}} - U)\beta_m$$

is integral for integral U.

Again, if we assume

$$P \mid E_h' \qquad \text{for } \deg P < h, \tag{3.11}$$

it is clear that (3.10) reduces to

$$\beta_m = G_m - (-1)^{h-1} e E_h^{d+1} \sum_{\deg P = h} \frac{1}{P}, \tag{3.12}$$

provided

$$\mu(m) = nh(p-1), \qquad p^{nh} - 1 \mid m;$$

in (3.12) the integers e and d are determined as above—with $s = 1$. (Actually (3.12) follows from

$$P \mid E_h' \qquad \text{for } \deg P \mid h, \deg P \neq h.)$$

In particular, (3.11) will be satisfied if $L_{h-1} \mid E_h'$, that is, if

$$L_{h-1}^2 \mid E_h'. \tag{3.13}$$

Clearly (1.4) implies (3.13)—except when $p^n = 2 = h$.

Finally, if (3.11) be replaced by

$$P \mid E_h' \qquad \text{for } \deg P \mid h, \tag{3.14}$$

it is evident that (3.12) reduces to $\beta_m = G_m$. Now (3.14) is equivalent to $x^{p^{nh}} - x \mid E_h'$, or what is the same,

$$L_h \mid E_h. \tag{3.15}$$

Thus we have the result that *if* (3.15) *holds, then* β_m *is integral.* We remark that if (3.15) is satisfied, then

$$\lambda(u) = \sum_{h=0}^{\infty} \frac{L_h}{F_h} D_h u^{p^{nh}} = \sum_h \frac{D_h u^{p^{nh}}}{g(p^{nh} - 1)},$$

where D_h is integral, and therefore $\lambda(u) = u\varphi(u)$, where $\varphi(u) = 1 + \varphi_1(u)$, and $\varphi_1(u)$ is of the form (1.1). As a consequence, the coefficients in $u/\lambda(u)$ are also integral.

The last result may be carried a bit further. A series of the form $\varphi(u) = 1 + \varphi_1(u)$ evidently has the property that its reciprocal

$$(1 + \varphi_1)^{-1} = 1 - \varphi_1 + \varphi_1^2 - \cdots$$

is also of the same type. Accordingly, we call the series

$$1 + \sum_{m=1}^{\infty} A_m \frac{u^m}{g_m}, \tag{3.16}$$

where A_m is integral, a unit series. Now it is easy to show that the quotient $\lambda(u)/u$ is an integral Hurwitz series (and therefore a unit series) if and only if condition (3.15) holds. But we have just seen that this condition also implies that β_m is integral. This in turn implies that $f(u)/u$ is a unit series, and therefore

$$L_h \mid A_h. \tag{3.17}$$

Thus we have the result that (3.15) and (3.17) are equivalent. Another way of putting it is that *the quotient* $f(u)/u$ *is a unit series if and only if* $\lambda(u)/u$ *is a unit series.*

4. **The general series.** We shall use the following notation. Put

$$f(u) = \sum_{m=1}^{\infty} \alpha_m \frac{u^m}{g_m} \qquad (\alpha_1 = 1); \tag{4.1}$$

let

$$\lambda(u) = \sum_{m=1}^{\infty} \epsilon_m \frac{u^m}{g_m} \qquad (\epsilon_1 = 1) \tag{4.2}$$

denote the inverse of $f(u)$. We shall also write $\alpha(m)$ and $\epsilon(m)$ in place of α_m and ϵ_m, respectively. To begin with, we require the following lemma which is independent of any hypothesis on $f(u)$—except that α_m is integral.

LEMMA 4. *For P irreducible of degree k,*

$$\frac{1}{F_k}\left(\sum_1^{\infty} \frac{\alpha_m}{g_m} u^m\right)^{p^{nk}} \equiv \sum_{i=1}^{\infty} \frac{\alpha(p^{nk(i-1)})}{F_{ki}} u^{p^{nki}} \pmod{P}. \tag{4.3}$$

To prove this formula, note that the left member of (4.3)

$$= \sum_{m=1}^{\infty} \alpha_m^{p^{nk}} \frac{g(p^{nk}m)}{F_k g^{p^{nk}}(m)} \frac{u^{p^{nk}m}}{g(p^{nk}m)}, \tag{4.4}$$

and the coefficient [1; Theorem 2]

$$\frac{g(p^{nk}m)}{F_k g^{p^{nk}}(m)} \equiv 0 \qquad \text{for } m \neq p^{ns},$$

while if $m = p^{ns}$, it reduces to

$$\frac{F_{k+s}}{F_k F^{p^{nk}}} = \begin{Bmatrix} k+s \\ k \end{Bmatrix} \equiv \begin{cases} 0 \text{ for } k \nmid s, \\ 1 \text{ for } k \mid s, \end{cases}$$

as previously observed. Therefore, (4.4) becomes

$$\sum_{s=0}^{\infty} \alpha(p^{ns}) \frac{u^{p^{nk(s+1)}}}{F_{k(s+1)}} \pmod{P},$$

so that we have proved (4.3).

We now introduce the hypothesis

$$P \mid \epsilon_m \qquad \text{for } p^{nk} < m. \tag{4.5}$$

Then if (4.5) holds, $\lambda(f) = u$ implies

$$u \equiv \sum_{m=1}^{p^{nk}} \frac{\epsilon_m}{g_m} f^m \pmod{P}. \tag{4.6}$$

Now let

$$f_0 = u - \frac{\epsilon(p^{nk})}{F_k} f^{p^{nk}}, \tag{4.7}$$

so that by (4.6)

$$f_0 \equiv f + \sum_{m=2}^{p^{nk}-1} \frac{\epsilon_m}{g_m} f^m;$$

therefore, by a familiar principle we have

$$f^{p^{nk}-1} \equiv f_0^{p^{nk}-1}. \tag{4.8}$$

Next, substituting from (4.3) in (4.7) we get

$$f_0 \equiv u - \epsilon(p^{nk}) \sum_{i=1}^{\infty} \alpha(p^{nk(i-1)}) \frac{u^{p^{nki}}}{F_{ki}}. \tag{4.9}$$

To further simplify the right member of (4.9) we require certain congruences involving $\alpha(p^{nk})$ and $\epsilon(p^{nk})$. Returning to (4.6) we pick out the coefficient of $u^{p^{nk}}/g(p^{nk})$ on the right side. It is not difficult to see that the only terms contributing to this coefficient are those for $m = 1$ or p^{nk}; this gives

$$\alpha(p^{nk}) + \epsilon(p^{nk}) \equiv 0. \tag{4.10}$$

More generally if we examine the coefficient of $u^{p^{nki}}/g(p^{nki})$ we get in exactly the same way

$$\alpha(p^{nki}) + \frac{F_{ki}}{F_k F_{k(i-1)}^{p^{nk}}} \epsilon(p^{nk}) \alpha(p^{nk(i-1)}) \equiv 0,$$

from which follows

$$\alpha(p^{nki}) \equiv (-1)^i \epsilon^i(p^{nk}) \equiv \alpha^i(p^{nk}). \tag{4.11}$$

Substituting in (4.9) we get finally

$$f_0 \equiv \sum_{i=0}^{\infty} \alpha(p^{nki}) \frac{u^{p^{nki}}}{F_{ki}}. \tag{4.12}$$

We have therefore proved

LEMMA 5. *If P is irreducible of degree k, and (4.5) is satisfied, then*

$$\left\{\sum_{1}^{\infty} \alpha_m \frac{u^m}{g_m}\right\}^{p^{nk}-1} \equiv \left\{\sum_{i=0}^{\infty} \alpha(p^{nki}) \frac{u^{p^{nki}}}{F_{ki}}\right\}^{p^{nk}-1} \pmod{P}. \tag{4.13}$$

This result may be compared with (2.12), the corresponding result in the linear case. It is then clear that the proof of Lemma 2 carries over without any change, and thus we get

LEMMA 6. *If the hypothesis of Lemma 5 is satisfied then*

$$\frac{1}{g(p^{nks}-1)} f^{p^{nks}-1} \equiv \frac{1}{g(p^{nk}-1)} \left\{\sum_{i=s-1}^{\infty} \alpha(p^{nki}) \frac{u^{p^{nki}}}{F_{ki}}\right\}^{p^{nk}-1} \times \frac{u^{p^{nk(s-1)}-1}}{g(p^{nk(s-1)}-1)}. \tag{4.14}$$

The remainder of the discussion in §2 can also be carried over without difficulty to the general case. We may state

LEMMA 7. *If* (4.5) *holds and* $k \mid h$, *then*

$$\frac{1}{g(p^{nh}-1)} f^{p^{nh}-1} \equiv \sum_m \gamma_m \frac{u^m}{g_m} \pmod{P}, \tag{4.15}$$

where $\gamma_m = 0$ *unless*

$$\mu(m) = nh(p-1), \qquad h = ks, \qquad p^{nk}-1 \mid m, \qquad p^{nk(s-1)} \mid m+1 \tag{4.16}$$

hold, while if (4.16) *holds then*

$$\gamma_m \equiv \frac{(-1)^{nk}}{\prod b_i!}\, \alpha^d(p^{nh}), \tag{4.17}$$

where

$$m - p^{nk(s-1)} + 1 = \sum_i b_i p^i \qquad (0 \le b_i < p), \qquad d = \sum_{i,j} i p^j b_{nki+j}\,.$$

We may now derive the main theorem. Clearly

$$\frac{u}{f} = \frac{\lambda(f)}{f} = \sum_{i=1}^{\infty} \epsilon_i \frac{f^{i-1}}{g_i} = \sum_{i=1}^{\infty} \frac{g_{i-1}}{g_i}\, \epsilon_i \frac{f^{i-1}}{g_{i-1}}.$$

Put

$$\frac{u}{f} = \sum_{m=0}^{\infty} \beta_m \frac{u^m}{g_m}, \qquad \frac{f^i}{g_i} = \sum_{m=i}^{\infty} \gamma_m^{(i)} \frac{u^m}{g_m},$$

where $\gamma_m^{(i)}$ is integral; then

$$\beta_m = \sum_{i \le m+1} \frac{g_{i-1}}{g_i}\, \epsilon_i \gamma_m^{(i-1)}. \tag{4.18}$$

Now it is easy to see that

$$\frac{g_i}{g_{i-1}} = L_h \qquad (p^{nh} \mid i,\ p^{n(h+1)} \nmid i). \tag{4.19}$$

Hence, if we assume the hypothesis

$$L_j \mid \epsilon_i \qquad \text{for } p^{nj} < i, \tag{4.20}$$

it is evident from (4.19) that $g_{i-1}\epsilon_i/g_i$ is integral unless $i = p^{nh}$. Thus (4.18) reduces to

$$\beta_m = G_m + \sum_h \frac{\epsilon(p^{nh})}{L_h} \gamma_m^{(p^{nh}-1)},$$

where G_m is integral. Note that for $i = p^{nh}$ condition (4.20) becomes $L_{h-1} \mid \epsilon(p^{nh})$. Then by Lemma 7, if $\mu(m) = nh(p-1)$, we have

$$\beta_m = G_m + \frac{\epsilon(p^{nh})}{L_h}\gamma_m^{(p^{nh}-1)}.$$

Hence, comparing with the corresponding point in §3 it is clear how we may complete the proof of

THEOREM 2. *Suppose the Hurwitz series* (4.1) *has the inverse* (4.2) *which satisfies the condition*

$$L_i \mid \epsilon_m \qquad \text{for } p^{ni} < m.$$

Define β_m *by*

$$\frac{u}{f(u)} = \sum_{m=0}^{\infty} \beta_m \frac{u^m}{g_m} \qquad (\beta_0 = 1),$$

and $\mu(m)$ *as in* (2.20). *If the system*

(4.21) $\mu(m) = nh(p-1), \quad h = ks, \quad p^{nh} - 1 \mid m, \quad p^{nh(s-1)} \mid m+1$

is inconsistent then β_m *is integral, while if* (4.21) *is consistent then*

$$\beta_m = G_m - e\epsilon'(p^{nh}) \sum_{k \mid h} \epsilon^d(p^{nk}) \sum_{\deg P = k} \frac{P'}{P}, \tag{4.22}$$

where $\epsilon(p^{nh}) = L_{h-1}\epsilon'(p^{nh})$, *and* d *and* e *are defined as in Theorem* 1. (*The outer summation will extend over all* k *dividing* h *only when* $m = p^{nh} - 1$.)

In place of (4.22) we may also write

$$\beta_m = G_m - e\epsilon'(p^{nh}) \sum_{k} (-1)^{k-1} \epsilon^d(p^{nk}) L_{k-1} \sum_{P} \frac{1}{P}. \tag{4.23}$$

As in the case of Theorem 1, there are a number of immediate corollaries of Theorem 2. For example

$$\beta_{m,U} = U(U^m - 1)\beta_m$$

is integral for U an arbitrary polynomial; if $\mu(m) = nh(p-1)$, then it follows as before that $(U^{p^{nh}} - U)\beta_m$ is integral. Again if we assume

$$P \mid \epsilon'(p^{nh}) \qquad \text{for } \deg P < h, \tag{4.24}$$

then (4.23) becomes

$$\beta_m = G_m - (-1)^{h-1} e\epsilon^{d+1}(p^{nh}) \sum_{\deg P = h} \frac{1}{P},$$

provided $\mu(m) = nh(p-1)$, $p^{nh} - 1 \mid m$; otherwise $\beta_m = G_m$; in particular, (4.23) will hold if (4.20) is replaced by

$$L_i^2 \mid \epsilon_m \qquad \text{for } p^{ni} < m.$$

In the next place, if (4.24) be replaced by

$$P \mid \epsilon'(p^{n^h}) \qquad \text{for } \deg P \mid h, \tag{4.25}$$

then (4.23) reduces to $\beta_m = G_m$. Now (4.20) and (4.25) are together equivalent to

$$L_i \mid \epsilon_m \qquad \text{for } p^{n^i} \leq m. \tag{4.26}$$

But if (4.26) is satisfied, then the inverse function (4.2) becomes

$$\lambda(u) = \sum_{m=1}^{\infty} \frac{L_i \delta_m}{g_m} u^m = u \sum_{m=1}^{\infty} \frac{L_i \delta_m g_{m-1}}{g_m} \frac{u^m}{g_{m-1}},$$

where $L_i \delta_m = \epsilon_m$. Clearly, by (4.19), $L_i g_{m-1}/g_m$ is integral, and therefore $\lambda(u) = u\varphi(u)$, where $\varphi(u)$ is a Hurwitz series. Thus in the present case also *if* (4.26) *holds, then the coefficients of both* u/f *and* u/λ *are integral.* It is not clear, however, whether the final result in §3 can be extended to the general case.

Bibliography

1. L. Carlitz, *An analogue of the von Staudt-Clausen theorem*, this Journal, vol. 3(1937), pp. 503–517.
2. L. Carlitz, *The reciprocal of certain series*, this Journal, vol. 9(1942), pp. 234–243.

Duke University.

THE POISSON INTEGRAL REPRESENTATION OF FUNCTIONS WHICH ARE POSITIVE AND HARMONIC IN A HALF-PLANE

By L. H. Loomis and D. V. Widder

A theorem of Herglotz states that a function is positive and harmonic in a circle if and only if it can be expressed as a Poisson-Stieltjes integral with non-decreasing integrator function (see, for example, [1; 571] for a proof and for a reference to the original paper). A corresponding result was stated and proved by S. Verblunsky [2] for a half-plane. The proof was obtained by transformation of the half-plane into a circle. We propose to give here an independent proof without use of any such transformation. We obtain incidentally an interesting uniqueness result, Theorem 3, which seems not to have been stated explicitly before. The proof assumes no more recondite knowledge of a harmonic function than that it cannot have a minimum inside its region of definition.

Theorem 1. *If* $\varphi(x)/(1+x^2) \in L$ *in* $(-\infty, \infty)$ *and is continuous at* x_0, *then*

$$F(x, y) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{y}{y^2 + (t-x)^2}\,\varphi(t)\,dt \tag{1}$$

is harmonic for $y > 0$ *and* $F(x_0, 0+) = \varphi(x_0)$.

As $|t| \to \infty$, the expression $(t^2+1)/(t-x)^2$ approaches 1, and so has a finite upper bound $M_\delta(x)$ over $|t-x| \geq \delta$. Hence for $y \neq 0$

$$\int_{x+\delta}^{\infty} \frac{y\varphi(t)}{y^2+(t-x)^2}\,dt \ll yM_\delta(x)\int_{x+\delta}^{\infty}\frac{|\varphi(t)|}{1+t^2}\,dt.$$

It is thus clear that the integral (1) converges for $y > 0$. It represents a harmonic function there since the integrand is harmonic for each fixed t. Since

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{y}{y^2+(x-t)^2}\,dt = 1,$$

we also have, for any $\delta > 0$,

$$|F(x_0, y) - \varphi(x_0)| \leq M_\delta(x_0)\frac{y}{\pi}\left(\int_{-\infty}^{x_0-\delta} + \int_{x_0+\delta}^{\infty}\right)\frac{|\varphi(t)-\varphi(x_0)|}{1+t^2}\,dt + \underset{|t-x_0|\leq\delta}{\text{u.b.}}\,|\varphi(t)-\varphi(x_0)|.$$

Hence

$$\overline{\lim_{y\to 0+}}\,|F(x_0, y) - \varphi(x_0)| \leq \underset{|t-x_0|\leq\delta}{\text{u.b.}}\,|\varphi(t)-\varphi(x_0)|.$$

Since the right side approaches zero with δ, the theorem is proved.

Received May 11, 1942.

THEOREM 2. *If $f(x, y)$ is harmonic and non-negative for $y > 0$, then $f(x, y)/(1 + x^2) \epsilon L$ in $(-\infty, \infty)$ for each positive y.*

Consider the function

$$F_{R,\delta}(x, y) = \frac{1}{\pi} \int_{-R}^{R} \frac{yf(t, \delta)}{y^2 + (t - x)^2} dt \qquad (\delta > 0).$$

By Theorem 1 it tends to $f(x, \delta)$ or to zero according as $| x | < R$ or $| x | > R$. That is, the harmonic function $f(x, y + \delta) - F_{R,\delta}(x, y)$ is non-negative for $y = 0$. It has the same property for $y > 0$. For, if it were negative at a point with positive ordinate, we would enclose this point in a square with sides $x = \rho$, $x = -\rho$, $y = 2\rho$, $y = 0$. On the first two sides

$$| F_{R,\delta}(x, y) | \leq \frac{\rho}{\pi(\rho - R)^2} \int_{-R}^{R} f(t, \delta)\, dt \qquad (\rho > R),$$

and on the third

$$| F_{R,\delta}(x, y) | \leq \frac{1}{2\pi\rho} \int_{-R}^{R} f(t, \delta)\, dt.$$

That is, $f(x, y + \delta) - F_{R,\delta}(x, y)$ would, for large ρ, be smaller at the point where it was assumed negative than on the boundary of the square. This is impossible, so that for $y > 0$

$$(3) \qquad \lim_{R \to \infty} F_{R,\delta}(x, y) = F_{\delta}(x, y) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{yf(t, \delta)}{y^2 + (t - x)^2} dt \leq f(x, y + \delta).$$

Setting $x = 0$, $y = 1$, we have the desired result.

THEOREM 3. *If $f(x, y)$ is harmonic and non-negative for $y \geq 0$, vanishing for $y = 0$, then it is a positive constant multiple of y.*

By the Schwarz reflection principle $f(x, y)$ is harmonic everywhere if we set $f(x, -y) = -f(x, y)$. Hence, in the Fourier expansion for $f(r \cos \theta, r \sin \theta) = f(re^{i\theta})$ the cosine terms vanish and

$$f(re^{i\theta}) = \sum_{n=1}^{\infty} b_n r^n \sin n\theta,$$

$$r^n b_n = \frac{2}{\pi} \int_0^{\pi} f(re^{i\theta}) \sin n\theta\, d\theta.$$

But then, since $| \sin n\theta | \leq n | \sin \theta |$ we have

$$| r^n b_n | \leq \frac{2n}{\pi} \int_0^{\pi} f(re^{i\theta}) \sin \theta\, d\theta = nrb_1.$$

This holds for all $r > 0$, so that $b_n = 0$ when $n > 1$. This completes the proof.

THEOREM 4. *The function $f(x, y)$ is harmonic and non-negative for $y > 0$ if and only if*

$$f(x, y) = Ay + \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{y}{y^2 + (t - x)^2}\, d\alpha(t), \tag{4}$$

where $\alpha(t)$ is non-decreasing and A is a positive constant.

The sufficiency of the condition is evident. To prove the necessity, observe that the function $F_\delta(x, y)$ defined in the proof of Theorem 2 approaches $f(x, \delta)$ when y approaches zero, by Theorems 1 and 2. By inequality (3) and Theorem 3,

$$f(x, y + \delta) - F_\delta(x, y) = A_\delta y \qquad (A_\delta > 0, y > 0),$$

$$f(x, y) = \lim_{\delta \to 0+} \left(A_\delta y + \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{y(1 + t^2)}{y^2 + (t - x)^2}\, d\beta_\delta(t) \right),$$

$$\beta_\delta(x) = \int_{-\infty}^{x} \frac{f(t, \delta)}{1 + t^2}\, dt \leq f(0, 1 + \delta).$$

Since $\beta_\delta(x)$ is non-decreasing and uniformly bounded for $-\infty < x < \infty$ and $0 < \delta \leq 1$, we may apply the Helly theorem [3] and the Helly-Bray theorem [3] in the standard way to obtain

$$f(x, y) = Ay + \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{y(1 + t^2)}{y^2 + (t - x)^2}\, d\beta(t). \tag{5}$$

(These theorems do not apply to the infinite interval unless a jump at infinity in the integrator function is admitted. We admit this jump in the first instance, then remove it, absorbing the resulting constant in A.) Here $\beta(t)$ is non-decreasing and bounded since it is the limit of a suitable sequence of functions taken from the set $\beta_\delta(t)$. Equation (5) is equivalent to equation (4) if we set

$$\alpha(x) = \int_0^x (1 + t^2)\, d\beta(t) \qquad (-\infty < x < \infty).$$

BIBLIOGRAPHY

1. M. H. STONE, *Linear Transformations in Hilbert Space and their Applications to Analysis*, American Mathematical Society Colloquium Publications, vol. XV(1932).
2. S. VERBLUNSKY, *On positive harmonic functions in a half-plane*, Proceedings of the Cambridge Philosophical Society, vol. 31(1935), pp. 482–507.
3. D. V. WIDDER, *The Laplace Transform*, Princeton University Press, 1941, pp. 26–32.

HARVARD UNIVERSITY.

PERIODICALS PUBLISHED BY DUKE UNIVERSITY

American Literature. A quarterly journal devoted to research in American Literature, published with the cooperation of the American Literature Group of the Modern Language Association of America. Subscription, $4.00 per year. Back volumes, $5.00 each.

Character and Personality. A psychological quarterly devoted to studies of behavior and personality. Subscription, $2.00 per year. The first number was published September, 1932.

Contributions to Psychological Theory. A monograph series dealing with problems of psychological theory in the widest sense, including their relations to other fields of inquiry. The monographs appear irregularly. Subscription, $5.00 per volume of approximately 450 pages.

Duke Mathematical Journal.

Ecological Monographs. A quarterly journal devoted to the publication of original researches of ecological interest from the entire field of biological science. Subscription, $6.00 per year. The first number was published January, 1931.

Hispanic American Historical Review. A quarterly review dealing with the history of the Latin-American countries. Subscription, $4.00 per year.

Law and Contemporary Problems. A quarterly published by the School of Law, presenting in each issue a symposium on a problem of current importance having significant legal aspects. Subscription, $2.00 per year. The first number was published September, 1933.

The South Atlantic Quarterly. A magazine of modern opinion and discussion, founded in 1902. Subscription, $3.00 per year.

The Southern Association Quarterly. As official organ of the Southern Association of Colleges and Secondary Schools, it contains the proceedings of the annual meeting, together with much additional material directly related to the work of the Association. Subscription, $4.00 per year.

DUKE UNIVERSITY PRESS
DURHAM, NORTH CAROLINA

CONTENTS

Non-analytic class field theory and Grunwald's theorem.
By George Whaples 455
n-to-one mappings of linear graphs. By Paul W. Gilbert 475
Monotone transformations. By A. D. Wallace 487
Normal bases of cyclic fields of prime-power degree. By Sam Perlis 507
Completely monotone functions in partially ordered spaces. By S. Bochner 519
Convergence in length and convergence in area.
By T. Radó and P. Reichelderfer 527
The absolute Cesàro summability of trigonometrical series.
By Fu Traing Wang 547
Theory of equivalence relations. By Oystein Ore 573
The reciprocal of certain types of Hurwitz series. By L. Carlitz 629
The Poisson integral representation of functions which are positive and harmonic in a half-plane. By L. H. Loomis and D. V. Widder 643